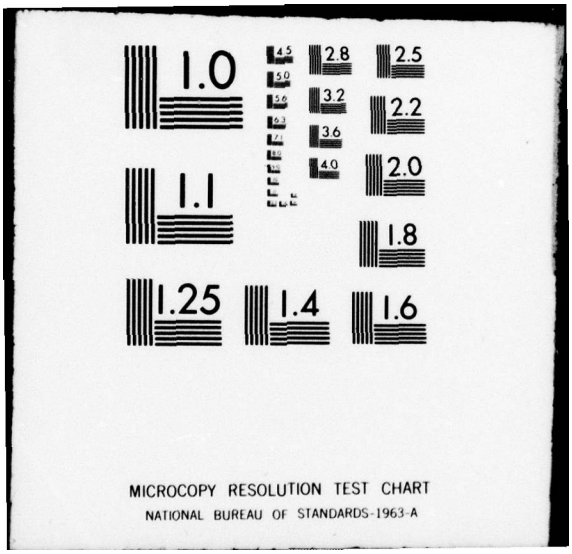
AD-A064 991 OFFICE OF NAVAL RESEARCH ARLINGTON VA F/G 15/5
NAVAL RESEARCH LOGISTICS QUARTERLY. VOLUME 25, NUMBER 3.(U)
SEP 78

UNCLASSIFIED NL

1 OF 3
AD
A064991


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

NAVAL RESEARCH LOGISTICS QUARTERLY.

Volume 25, Number 3.

SEPTEMBER 1978
VOL. 25, NO. 3

OFFICE OF NAVAL RESEARCH

79 02 12 080

NAVSO P-1278

NAVAL RESEARCH LOGISTICS QUARTERLY

The Naval Research Logistics Quarterly is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Information for Contributors is indicated on inside back cover.

The Naval Research Logistics Quarterly is published by the Office of Naval Research in the months of March, June, September, and December and can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Subscription Price: $11.15 a year in the U.S. and Canada, $13.95 elsewhere. Cost of individual issues may be obtained from the Superintendent of Documents.

Issuance of this periodical approved in accordance with Department of the Navy Publications and Printing Regulations P-35 (Revised 1-74).

SOME APPROXIMATIONS IN MULTI-ITEM, MULTI-ECHELON* INVENTORY SYSTEMS FOR RECOVERABLE ITEMS

John A. Muckstadt

Cornell University
Ithaca, New York

ABSTRACT

The optimization problem as formulated in the METRIC model takes the form of minimizing the expected number of total system backorders in a two-echelon inventory system subject to a budget constraint. The system contains recoverable items — items subject to repair when they fail. To solve this problem, one needs to find the optimal Lagrangian multiplier associated with the given budget constraint.

For any large-scale inventory system, this task is computationally not trivial. Fox and Landi proposed one method that was a significant improvement over the original METRIC algorithm. In this report we first develop a method for estimating the value of the optimal Lagrangian multiplier used in the Fox-Landi algorithm, present alternative ways for determining stock levels, and compare these proposed approaches with the Fox-Landi algorithm, using two hypothetical inventory systems — one having 3 bases and 75 items, the other 5 bases and 125 items. The comparison shows that the computational time can be reduced by nearly 50 percent.

Another factor that contributes to the higher requirement for computational time in obtaining the solution to two-echelon inventory systems is that it has to allocate stock optimally to the depot as well as to bases for a given total-system stock level. This essentially requires the evaluation of every possible combination of depot and base stock levels — a time-consuming process for many practical inventory problems with a sizable system stock level. This report also suggests a simple approximation method for estimating the optimal depot stock level. When this method was applied to the same two hypothetical inventory systems indicated above, it was found that the estimate of optimal depot stock is quite close to the optimal value in all cases. Furthermore, the increase in expected system backorders using the estimated depot stock levels rather than the optimal levels is generally small.

I. INTRODUCTION

Almost a decade ago, Sherbrooke formulated the well-known METRIC model for determining optimal stock levels for recoverable items — items subject to repair when they fail — in a two-echelon setting [3]. Briefly, the two-echelon system consists of several locations, called bases, at which primary demands occur; these bases are in turn resupplied as necessary by a central repair and inventory-stocking point called a depot. When a failure occurs at a base, a demand is placed on base supply for a corresponding replacement part. The failed part is then either repaired at that base, or is sent to the depot for repair, depending on the nature of the

*This research was partially supported by the Office of Naval Research under Contract N00014-75-C-1172, Task NR 042-335, and by the RAND Corporation.

failure. Resupply of base supply comes from the base maintenance organization if repair is accomplished at the base; otherwise, resupply comes from the depot. In either case, the organization resupplying the base supply activity does so by exchanging a serviceable part for the failed part. Thus the inventory policy for placing orders on the base's maintenance organization or the depot is an $(s-1,s)$ policy.

Sherbrooke presented a model that can be used to determine both depot and base stock levels for all items for this system. In particular, the problem he formulated was that of minimizing the expected total number of base back orders existing at an arbitrary time subject to a constraint on system investment, that is,

$$\min \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{x>s_{ij}} (x - s_{ij}) p[x|\lambda_{ij} T_{ij}(s_{i0})]$$

$$\text{(P1)} \qquad \text{subject to } \sum_{j=0}^{m} \sum_{i=1}^{n} c_i s_{ij} \leqslant C,$$

where

- n = the number of items,
- m = the number of bases,
- s_{ij} = the stock level at base j for item i, a non-negative integer,
- s_{i0} = the depot stock level for item i, a non-negative integer,
- λ_{ij} = the expected daily demand rate for item i at base j,
- c_i = the unit cost for item i,
- C = the budget constraint,
- $T_{ij}(s_{i0})$ = the average resupply time for base j for item i, given that the depot stock level for item i is s_{i0} and
- $p(x|y)$ = the probability that x units are in the resupply system when the expected number of units in the resupply system is y.

Furthermore, Sherbrooke shows that $T_{ij}(s_{i0})$ can be expressed as

$$T_{ij}(s_{i0}) = r_{ij}A_{ij} + (1 - r_{ij})[B_{ij} + \delta(s_{i0}) \cdot D_i],$$

where

- A_{ij} = the average base repair time for item i at base j,
- r_{ij} = the proportion of demands requiring base repair for item i at base j,
- B_{ij} = the average order-and-ship time at base j for item i,
- D_i = the average depot repair-cycle time for item i,
- $\delta(s_{i0}) \cdot D_i$ = expected depot backorders/expected depot daily demand rate $= (1/\lambda_i) \sum_{x>s_{i0}} (x - s_{i0}) p(x|\lambda_i D_i)$, the expected delay per depot demand for item i, and
- λ_i = $\sum_{j=1}^{m} (1 - r_{ij})\lambda_{ij}$, the expected daily depot demand for item i.

In the remainder of the report, i will refer to an item and j will refer to a base ($j = 0$ represents the depot). For a complete description of the problem's background and formulation, see Ref. [3].

Subsequently, Fox and Landi suggested a Lagrangian approach for solving problem P1 [2]. One obstacle to the implementation of METRIC using the Fox-Landi algorithm is the require-

ment of estimating an appropriate value for the Lagrangian multiplier. Another important and related problem is the lengthy computer run time required to obtain an optimal solution to problem P1 when using their algorithm. A large portion of this computational effort is related to searching for the optimal depot stock level. This search is particularly time-consuming for items having a high average number of units in the depot repair cycle since the amount of computation required by their algorithm is roughly proportional to the number of depot stock levels explicitly examined.

In this paper, we first develop an approach for obtaining an estimate of the optimal Lagrange multiplier value required in the Fox-Landi algorithm, present two new methods for determining stock levels, and compare these methods with the Fox-Landi method and other techniques. The proposed approach eliminates the particularly time-consuming portion of the Fox-Landi algorithm devoted to searching for the best Lagrange-multiplier value and significantly reduces computation time for determining stock levels without degrading the quality of the solution.

We then present a method for estimating the optimal depot stock level. Limited computational experience indicates that this method is easy to implement, provides a very good estimate of the optimal depot stock level, and is particularly useful for items having a high average number of units in the depot repair cycle. For these items it is possible to reduce computation time required by the Fox-Landi algorithm by as much as 90 percent.

II. THE APPROXIMATION PROBLEM

In this section, we first construct a problem that is a continuous approximation to problem P1. We then state and prove two theorems that are the basis for an algorithm that can be used to solve this approximating problem.

Two useful probability distributions for describing the demand process are the Poisson and negative-binomial distributions. As shown in Ref. [1], this implies that if demand has a Poisson or a negative-binomial distribution, then, for a given value of $\lambda_{ij}T_{ij}(s_{i0})$, the probability distribution representing the number of units in resupply of items i at base j at any time, $p[x|\lambda_{ij}T_{ij}(s_{i0})]$, is a Poisson or a negative-binomial distribution, respectively.

Experimental data gathered during the conduct of this study indicate that, when $p[x|\lambda_{ij}T_{ij}(s_{i0})]$ is either a Poisson or a negative-binomial distribution, the expected number of back orders at each location can be closely approximated by an exponential function. This is not unexpected. First, for budgets of practical interest, the item stock levels, s_{ij}, are normally much larger than the average demand during the resupply time. In fact, the probability of running out of stock during the resupply time is often much less than 0.15 in real applications. Thus, the only probabilities entering the backorder calculation are the tail probabilities of the distribution. In the tails, both the Poisson and negative-binomial distributions behave almost like the geometric distribution; that is, each succeeding probability is roughly a constant proportion of its predecessor. Consequently, when s_{ij} is large relative to $\lambda_{ij}T_{ij}(s_{i0})$, the expected number of backorders existing at any time at location j for item i is approximately a geometric function of s_{ij}. Therefore, an exponential function is a useful continuous approximation to this relationship between expected backorders at a location and the item's stock level at that location.

Furthermore, total expected base backorders for each item exhibit this same behavior. If demand has either a Poisson or a negative-binomial distribution (or, for that matter, any compound Poisson distribution), then the total number of units of an item in resupply across all

bases has either a Poisson or a negative-binomial distribution, respectively, if we assume independence of demand and common variance-to-mean ratio among base demand distributions. Since, in most practical situations, total system stock substantially exceeds the total expected number of units in resupply, the tail of the distribution describing the total number of units in resupply is the only portion of the distribution of importance. As an approximation, this distribution can be used to determine the nature of the relationship between total expected base backorders and total system stock. For the reasons discussed previously, an exponential function should also adequately represent this relationship.

Thus we will approximate

$$\sum_{i=1}^{m} \sum_{x>s_{ij}} (x - s_{ij}) p[x|\lambda_{ij} T_{ij}(s_{i0})]$$

with the exponential function

$$B_i(N_i) \equiv a_i e^{-b_i N_i}.$$

In this approximation, N_i represents total system stock. In practice, the parameters $a_i > 0$ and $b_i > 0$ are estimated using regression analysis. The data used in the regression analysis are the backorder data obtained from the solution to the problem

$$\min \sum_{i=1}^{m} \sum_{x>s_{ij}} (x - s_{ij}) p[x|\lambda_{ij} T_{ij}(s_{i0})]$$

$$\text{subject to } \sum_{j=0}^{m} s_{ij} = N_i \quad \text{and}$$

$$s_{ij} = 0, 1, \ldots, N_i,$$

for several appropriate values of N_i.

We now formulate a continuous approximation to problem P1 in which the exponential approximation to total system backorders for an item is used. In this approximation problem, the decision variables are the total system stock, N_i, rather than the stock levels for each location, s_{ij}. This approximation problem is a vehicle for obtaining an estimate of the optimal Lagrangian-multiplier value used in the Fox-Landi algorithm. The approximation problem is formulated as problem P2:

$$\min \sum_{i=1}^{n} B_i(N_i)$$

(P2)
$$\text{subject to } \sum_{i=1}^{n} c_i N_i \leqslant C,$$

where

$$N_i \geqslant 0.$$

Note that N_i is a continuous variable in this approximation. The optimality (Kuhn-Tucker) conditions for this problem are as follows:

Find $\theta_1 \geqslant 0$ such that

(a)
$$\frac{dB_i}{dN_i} + \theta_1 c_i \geqslant 0,$$

(b) $$\sum_{i=1}^{n} c_i N_i \geq C,$$

$$N_i \geq 0,$$

(c) $$\theta_1 \left(\sum_{i=1}^{n} c_i N_i - C \right) = 0, \text{ and}$$

(d) $$N_i \left(\frac{dB_i}{dN_i} + \theta_1 c_i \right) = 0.$$

A relaxed version of problem P2 in which the nonnegativity constraint on the item stock level is removed is problem P3:

$$\min \sum_{i=1}^{n} B_i(N_i)$$

(P3) $$\text{subject to } \sum_{i=1}^{n} c_i N_i \leq C.$$

The optimality conditions for this problem are similar to those for P2. In condition (a) the inequality is replaced by an equality and the nonnegativity restriction on N_i is omitted. Also θ_2 will represent the Lagrangian multiplier for problem P3.

We now explore the relationship between problems P2 and P3 in detail.

Suppose we solve problem P3.* Let N_i^1 represent the optimal solution to problem P2, and N_i^2 represent the optimal solution to problem P3. If $N_i^2 \geq 0$ for all i, then $N_i^1 = N_i^2$ and the objective function values are equal.

If $N_i^2 < 0$ for at least one value of i, let

$$\bar{N}_i = \max(0,\ N_i^2)$$

and

$$\bar{C} \equiv \sum_{i=1}^{n} c_i \bar{N}_i > C.$$

Suppose problem P2 is modified slightly so that C is replaced by $\bar{C}$. This modified problem is called problem P4. The optimality conditions for this problem are the same as those for problem P2 after substituting $\bar{C}$ for C. Also, let $\bar{\theta}$ represent the optimal value of the Lagrangian multiplier for problem P4.

In solving problem P3, we will obtain a value for θ_2. It is easy to show that $\bar{\theta} = \theta_2$ and that $\bar{N}_i = \max(0,\ N_i^2)$ is an optimal solution to problem P4 by demonstrating that these values satisfy the Kuhn-Tucker conditions corresponding to problem P4. Consequently, the optimal solution to problem P4 is $N_i = \bar{N}_i = \max(0,\ N_i^2)$. Furthermore, the optimality conditions are satisfied when $\bar{\theta}$ is equal to θ_2.

THEOREM 1: $\theta_1 \geq \theta_2 = \bar{\theta}$.

*Section III develops the method for determining the solution to problem P3.

PROOF: The optimal objective function value for problem P2 is a convex, differentiable, strictly decreasing function of the available budget, C. Since the slope of this function is equal to the negative of the Lagrangian-multiplier value, $\theta_1 \geqslant \bar{\theta}$ since $C \leqslant \bar{C}$. But $\theta_2 = \bar{\theta}$, so $\theta_1 \geqslant \theta_2$.

Next we compare N_i^1 with $\bar{N}_i$. Suppose $\bar{C} > C$ so that $\theta_1 > \theta_2 = \bar{\theta}$. Let us examine the two cases $\bar{N}_i > 0$ and $\bar{N}_i = 0$ separately.

First, assume $\bar{N}_i > 0$. Then

$$\left.\frac{dB_i}{dN_i}\right|_{N_i=\bar{N}_i} + \bar{\theta}c_i = 0.$$

Furthermore, if $N_i^1 > 0$, then

$$\left.\frac{dB_i}{dN_i}\right|_{N_i=N_i^1} + \theta_1 c_i = 0.$$

Since

$$\theta_1 c_i > \bar{\theta}c_i = -\left.\frac{dB}{dN_i}\right|_{N_i=\bar{N}_i},$$

$$\left.\frac{dB_i}{dN_i}\right|_{N_i=\bar{N}_i} > \left.\frac{dB_i}{dN_i}\right|_{N_i=N_i^1},$$

and $N_i^1 < \bar{N}_i$. If $N_i^1 = 0$, then $\bar{N}_i > N_i^1$.

Next, assume $\bar{N}_i = 0$. Since

$$\left.\frac{dB_i}{dN_i}\right|_{N_i=0} + \theta_1 c_i > \left.\frac{dB_i}{dN_i}\right|_{N_i=0} + \bar{\theta}c_i \geqslant 0,$$

it follows that $N_i^1 = 0$ by complementary slackness. Thus we have proven the following theorem.

THEOREM 2. $\bar{N}_i \geqslant N_i^1$ and $\bar{N}_i > N_i^1$ *whenever* $\bar{N}_i > 0$.

Having established several important relationships among problems P2, P3, and P4, we next develop a simple algorithm for solving problem P2 and show how to find the solution to problem P3.

III. COMPUTING OPTIMAL SOLUTIONS FOR PROBLEMS P2 AND P3

Observe that the optimal solution to problem P3 must satisfy the two conditions

$$\frac{dB_i}{dN_i} + \theta_2 c_i = 0$$

and

$$\sum_{i=1}^{n} c_i N_i = C$$

because each $B_i(N_i)$ is a strictly decreasing function of N_i.

Since

$$B_i(N_i) = a_i e^{-b_i N_i},$$

where $a_i,\ b_i > 0$, the first condition states that

$$\theta_2 = \frac{a_i b_i e^{-b_i N_i}}{c_i} > 0.$$

Letting

$$\hat{\theta} \triangleq \ln \theta_2 = \ln \left(\frac{a_i b_i}{c_i} \right) - b_i N_i$$

and

$$d_i = \ln \left(\frac{a_i b_i}{c_i} \right),$$

we see that

$$N_i = \frac{d_i - \hat{\theta}}{b_i}.$$

From the second constraint we know that

$$\sum_{i=1}^{n} c_i \left(\frac{d_i - \hat{\theta}}{b_i} \right) = C.$$

Thus

$$\hat{\theta} = \frac{\sum_{i=1}^{n} (c_i d_i / b_i) - C}{\sum_{i=1}^{n} (c_i / b_i)}.$$

Letting

$$\alpha = \sum_{i=1}^{n} \frac{c_i d_i}{b_i} \quad \text{and} \quad \beta = \sum_{i=1}^{n} \frac{c_i}{b_i},$$

we can express $\hat{\theta}$ as

$$\hat{\theta} = \frac{\alpha - C}{\beta}.$$

Thus

(E1) $$\theta_2 = e^{(\alpha - C)/\beta}$$

and

(E2) $$N_i = \frac{d_i - \dfrac{\alpha - C}{\beta}}{b_i}.$$

Consequently, N_i is a *linear function of C.*

We may employ the following algorithm to find the optimal solution to problem P2. Let $I = \{1, \ldots, n\}$ and N_i^1 represent the optimal solution to problem P2.

STEP 0: Solve Problem P3 as described above, thereby obtaining an initial value for N_i, $i \in I$.

STEP 1: Set $N_i^1 = 0$ for all $N_i < 0$ during the last iteration and delete the corresponding i from I. Recompute α and β, where

$$\alpha = \sum_{i \in I} \left\{ \frac{c_i d_i}{b_i} \right\}$$

and

$$\beta = \sum_{i \in I} \left\{ \frac{c_i}{b_i} \right\}.$$

STEP 2: Using (E2), obtain new estimates of N_i for each $i \in I$. If $N_i \geqslant 0$ for all $i \in I$, then the optimal solution has been found, and $N_i^1 = N_i$ for all $i \in I$ and $N_i^1 = 0$ for all $i = 1, \ldots, n$ for which $i \notin I$. If there exists some i for which $N_i < 0$, return to Step 1.

It is clear that our solution satisfies all the optimality conditions for problem P2 except, possibly, condition (a) for $i \notin I$. However, at an earlier iteration (when i was deleted from I) we had

$$\left. \frac{dB_i}{dN_i} \right|_{N_i = \tilde{N}_i} + \tilde{\theta}_2 c_i = 0,$$

where $\tilde{\theta}_2$ and $\tilde{N}_i (<0)$ are earlier values of θ_2 and N_i, respectively. Since dB_i/dN_i is clearly increasing in N_i, and θ_2 increases at each iteration (Theorem 1 and its corollary), condition (a) must hold. Convergence is guaranteed since n is finite.

IV. A COMPARISON OF ALTERNATIVE SOLUTION PROCEDURES FOR SOLVING PROBLEM P1

In this section we review three algorithms for solving problem P1 and compare them to two algorithms designed to obtain a solution for problem P1 based on the solution to the approximating problem, problem P2.

The Sherbrooke Procedure

The first algorithm, a procedure originally proposed by Sherbrooke [3], is a marginal-analysis algorithm consisting of two phases. In the first phase, each item is examined independently. The optimization problem solved for item i in the first phase has the form:

$$Z_i(N_i) = \min \sum_{j=1}^{m} \sum_{x > s_{ij}} (x - s_{ij}) p[x | \lambda_{ij} T_{ij}(s_{i0})]$$

$$\text{(P5)} \qquad \text{subject to} \sum_{j=0}^{m} s_{ij} = N_i,$$

where

$$s_{ij} = 0, 1, \ldots,$$

and N_i is the total system stock available for distribution among the depot and bases. Problem P5 is solved by obtaining the solution to the $N_i + 1$ problems

$$\bar{Z}_i(N_i, s_{i0}) = \min \sum_{j=1}^{m} \sum_{x > s_{ij}} (x - s_{ij}) p[x|\lambda_{ij} T_{ij}(s_{i0})]$$

(P6) $$\text{subject to } \sum_{j=1}^{m} s_{ij} = N_i - s_{i0},$$

where

$$s_{ij} = 0, 1, \ldots,$$

and s_{i0} is fixed for $s_{i0} = 0, 1, \ldots, N_i$. Problem P6 can be solved via marginal analysis. Then

$$Z_i(N_i) = \min_{s_{i0}} \bar{Z}_i(N_i, s_{i0}),$$

where

$$s_{i0} = 0, \ldots, N_i.$$

The second-phase problem is

$$\min \sum_{i=1}^{n} Z_i(N_i)$$

$$\text{subject to } \sum_{i=1}^{n} c_i N_i \leqslant C,$$

where

$$N_i = 0, 1, \ldots.$$

Sherbrooke [3] suggests that a marginal-analysis algorithm be used to find a solution to this knapsack problem. Clearly other procedures could be employed to obtain an optimal solution. In any case, this approach requires a substantial amount of storage to save all the $Z_i(N_i)$ values. For moderate-sized problems having several thousand items, a storage requirement of 10^6 or more words may be needed to save these values. Furthermore, the computation time required to obtain these $Z_i(N_i)$ values for such problems is very large.

The Fox-Landi Procedure

Subsequently Fox and Landi [2] proposed a Lagrangian algorithm for solving problem P1. In particular, they formulated the relaxed version of problem P1 as problem P7:

(P7) $$\min \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{x > s_{ij}} (x - s_{ij}) p[x|\lambda_{ij} T_{ij}(s_{i0})] + \theta \sum_{j=0}^{m} \sum_{i=1}^{n} c_i s_{ij},$$

where

$$s_{ij} = 0, 1, \ldots,$$

and θ is the Lagrangian multiplier. Since problem P7 is separable by item, its optimal solution can be found by solving the n individual item problems

$$\min \sum_{j=1}^{m} \sum_{x > s_{ij}} (x - s_{ij}) p[x|\lambda_{ij} T_{ij}(s_{i0})] + \theta \sum_{j=0}^{m} c_i s_{ij}$$

subject to

$$s_{ij} = 0, 1, \ldots.$$

This problem, like problem P6 in Sherbrooke's two-phase method, is solved using a partitioning procedure, that is, it is reformulated as

$$\text{(P8)} \qquad \min_{s_{i0}=0,1,\ldots} \left\{ \theta c_i s_{i0} + \sum_{j=1}^{m} \min_{s_{ij}=0,1,\ldots} \left\{ \sum_{x>s_{ij}} (x - s_{ij}) p[x|\lambda_{ij} T_{ij}(s_{i0})] + \theta c_i s_{ij} \colon s_{i0} \text{ fixed} \right\} \right\},$$

or equivalently as

$$\text{(P9)} \qquad \min Z(s_{i0}; \theta)$$

where

$$s_{i0} = 0, 1, \ldots,$$

and

$$Z(s_{i0}; \theta) = \theta c_i s_{i0} + \sum_{j=1}^{m} \min_{s_{ij}} \left\{ \sum_{x>s_{ij}} (x - s_{ij}) p[x|\lambda_{ij} T_{ij}(s_{i0})] + \theta c_i s_{ij} \colon s_{ij} = 0, 1, \ldots; s_{i0} \text{ fixed} \right\}.$$

To determine $Z(s_{i0}; \theta)$, solve the m base problems

$$\min_{s_{ij}} \sum_{x>s_{ij}} (x - s_{ij}) p[x|\lambda_{ij} T_{ij}(s_{i0})] + \theta c_i s_{ij}.$$

The optimal s_{ij} is the smallest nonnegative integer for which

$$\sum_{x>s_{ij}} p[x|\lambda_{ij} T_{ij}(s_{i0})] \leqslant \theta c_i.$$

Problem P8 is solved for each item for a given value of θ. This yields a total investment cost corresponding to θ. In the Fox-Landi approach, the "optimal" value of θ is selected from a grid of M equally spaced values,

$$\theta_0 > \theta_1 > \ldots > \theta_M > 0.$$

The optimal value of θ is the θ_K, $K \in \{0, \ldots, M\}$, whose corresponding total investment cost is closest to C.

Fox and Landi suggest that their method is a single-pass method; that is, only one pass through the item data base is necessary to obtain the optimal solution. The storage requirement to effect this one-pass approach is potentially enormous. For a moderate-sized problem having 3000 items, 20 bases, and $M = 63$, almost 4 million item stock levels must be saved, plus possibly millions of additional item data elements reflecting fill rates, probability of no stockout at an arbitrary time, expected base backorders, and so on. Furthermore, because there may be no simple method for estimating suitable bounds on the values of the multipliers, much larger values of M may be required to ensure adequate approximation of the budget.

It has been the author's experience that Air Force personnel have difficulty estimating a reasonable range for θ for large problems. This is not surprising, because the data used in the model frequently change in real situations, thereby causing the optimal value of the multiplier to change. Furthermore, changing the multiplier's magnitude by 10^{-6} or less often causes the corresponding total cost to change by many millions of dollars. Consequently, 2^{10} values of θ

have been used in some Air Force applications to make the system "foolproof." In these cases, 60 million or more item stock levels would have to be explicitly stored, plus a considerable amount of other item and base data, to make the Fox-Landi algorithm truly a one-pass method.

On the other hand, if their method is altered so that the item data are passed through a second time, it is possible to eliminate virtually all of the requirement for secondary storage. In the first pass, only the running total cost corresponding to each θ_K, $K \in \{0, \ldots, M\}$, is saved. At the end of this phase, the "optimal" multiplier value, θ^*, is established. The second phase of the algorithm requires a second pass through the data base. In the second pass, the optimal stock levels for each location are found for all items by resolving problem P8 with $\theta = \theta^*$.

In some applications, the Fox-Landi one-pass method is clearly infeasible; that is, there may not be enough peripheral storage capacity to save all of the data. If storage capacity is available, there is a tradeoff between the time and cost required to store and access the data in the secondary memory using the one-pass method, and the time and cost to recompute the stock levels using the second method. For realistic Air Force problems, the two-pass method appears to be the only feasible approach, given current hardware constraints, if M is large enough to guarantee that a solution can be found that closely approximates the target budget.

The Bisection Method

A third way to solve problem P1 is a slight modification of the Fox-Landi algorithm called the bisection method, which employs a bisection search to find the optimal value for θ. This procedure requires initial upper and lower bounds on the optimal value of θ. Call these θ_U and θ_L, respectively. The bisection method is as follows:

STEP 1: Set $\bar{\theta} = (\theta_U + \theta_L)/2$.

STEP 2: Solve problem P8 with $\theta = \bar{\theta}$ for each item.

STEP 3: If the total cost of the solution obtained in Step 2 exceeds C, then replace θ_L with $\bar{\theta}$; otherwise, replace θ_U with $\bar{\theta}$.

STEP 4: If a stopping criterion has not been met (such as a fixed number of iterations or an error tolerance), return to Step 1; otherwise, stop.

The major drawback to the bisection approach is that a separate pass through the item data base is required at each iteration of the algorithm. This algorithm performs very well in terms of convergence, and we have found that it almost always produces solutions that are within 1/2 percent of the target budget using 10 bisections.

Comparison of Methods

The closeness of the solutions to the target budget generated by either the Fox-Landi method or the bisection algorithm depends on how broad a range of multiplier values must be searched for a fixed value of M or a fixed number of bisections. It should be pointed out that both of these methods only yield an approximation to the optimal multiplier value (assuming one exists).

Of the methods discussed thus far, it has been the experience of the author, as well as of Fox and Landi [2], that the latter two algorithms better Sherbrooke's algorithm in run times by an order of magnitude or more on real problems, given reasonable estimates of upper and lower

bounds for the Lagrangian multiplier. Thus, in the comparisons we will report, only these two Lagrangian methods will be discussed.

Approximation Methods

Earlier we described an approximation method for estimating the optimal value of θ and of each N_i. Several options are open for implementing this approximation method. One way to implement it is to use a two-phase approach. We call this approach the First Approximation Method. The values of a_i and b_i are computed in the first phase of this method, and the optimal value of θ is estimated using (E1). In the second phase, we solve problem P8 for each item, using the estimate of the optimal θ. This approach has two major advantages over the Fox-Landi method:

1. The estimate of the optimal multiplier can be obtained without prespecifying a range of values, and computation time to obtain the estimate does not depend on the uncertainty of the multiplier value.

2. The computation time to find an estimate of the optimal multiplier is much smaller.

If the two-pass version of the Fox-Landi algorithm is used, the second phase of that method and the second phase of the approximation method are the same. The one-pass version of the Fox-Landi algorithm requires considerably more storage, and also requires more computer time to determine the optimal stock levels, than this approximation method requires.

This approximation approach also has advantages over the bisection method:

1. Only two passes through the data base are required, as opposed to seven or more required for the bisection method in practice.

2. No stock levels need to be saved; in the bisection method it is necessary to save all stock levels and other data for three multiplier values.

Another algorithm can be employed that directly uses the results of the approximation problem, that is, problem P2. We call this approach the Second Approximation Method. This algorithm is of interest in situations in which only total system stock is computed for each item, and there is no interest in computing the optimal distribution of the assets. Determining the optimal allocation of a budget among items is of primary importance when purchasing inventory or making budgetary projections for spares for different systems. In these cases, distribution decisions are usually not very critical.

The Second Approximation algorithm also consists of two phases. In the first phase we estimate the values of the a_i and b_i parameters, and in the second phase we determine total system stock for each item, using the algorithm described in Section III, and rounding N_i to the nearest integer. The algorithm requires one pass through the item data base and one pass through an item file consisting of a_i, b_i, and c_i. The major advantage of this approach is that it eliminates the stock-allocation phase of both the Fox-Landi method and the First Approximation algorithm.

V. A COMPUTATIONAL COMPARISON OF VARIOUS ALGORITHMS

The Fox-Landi algorithm, the bisection algorithm, and the two approximation methods have been coded and tested on several sample sets of data for the Air Force's new F-15 fighter.

Since all of the tests yielded the same general results, we will discuss only two of them in detail. The first test had a 75-item sample and had 3 operating bases. The flying programs were very different at each base. In the second test, 125 items were included in the sample, with demands occurring at 5 bases; in this test, only the Fox-Landi and the two approximation methods were compared. The run times stated for both approximation algorithms include the time required to estimate the values of a_i and b_i. In all Fox-Landi calculations, a maximum of 128 multiplier values was examined; ten bisections were used in all applications of the bisection method. Furthermore, in both test cases all stock levels for all relevant multiplier values were stored in main memory. Thus, although the reported computation times, which include compile times, are roughly equal for all the algorithms, they are biased in favor of the Fox-Landi method because this type of storage would be impossible for larger problems. In addition, the range of multiplier values considered in the tests of the Fox-Landi and the bisection methods was selected after estimating the optimal multiplier value using the First Approximation Method. Thus the test results are biased in favor of them, since the range of multiplier values was much smaller than would normally be the case.

The data displayed in Tables 1 and 2 indicate how well each approach approximates a given target budget for the two test-data sets. Without a doubt, the bisection method produced solutions that best matched the target budgets, followed in order by the Second Approximation Method, the Fox-Landi method, and the First Approximation Method. As mentioned before, the results are biased in favor of both the Fox-Landi and the bisection methods due to the initialization of the range of multiplier values. From a practical viewpoint, all approaches worked acceptably well in meeting the target budgets. Furthermore, the stock levels generated by the various approaches were virtually the same for similar budgets. Consequently, total system expected backorders, for all practical purposes, are indistinguishable; that is, the backorder versus investment curves virtually coincide among these various approaches. Exact comparison of computed stock levels and expected backorders cannot be made among the competing methods since the allocation of the available budget in each case depends on the way each algorithm estimates the Lagrangian multiplier.

The area in which the methods clearly differ is in computation time. The approximation methods require substantially less time than either the Fox-Landi method or the more time-consuming bisection method. Other experimentation has shown that the percentage difference in computation times tends to be even greater as the number of items increases.

Thus, the approximation methods produce answers that are as good as those produced by the Fox-Landi method and the bisection method, but do it much more quickly than those methods. The bisection method does, however, match target budgets slightly better than the approximation methods. However, the approximation algorithms are virtually foolproof, which is perhaps their greatest advantage. The user does not have to specify the range of multiplier values or the number of bisections in advance. This eliminates one problem associated with implementing either the Fox-Landi or the bisection algorithm. In view of these observations, the approximation procedures developed here appear to be superior for use on real problems.

VI. ESTIMATION OF THE OPTIMAL DEPOT STOCK LEVEL

We have described Sherbrooke's algorithm and several Lagrangian methods for solving problem P1, and have demonstrated that it is possible to reduce significantly the computational requirement of the Fox-Landi method by solving an approximation problem to obtain a good estimate of an appropriate value for the Lagrangian multiplier. In this section we describe a different way to reduce the computational requirements of all the algorithms that have been discussed. As can be seen by reexamining Sherbrooke's approach (see problem P6) and the

TABLE 1 — *75-Item, 3-Base Test Case*

Target Budget ($ millions)	Total Cost ($ millions)			
	Bisection	Fox-Landi	First Approximation	Second Approximation
3.68	3.67	3.68	3.63	3.63
3.97	3.99	3.92	3.82	4.03
4.27	4.27	4.27	4.30	4.18
4.57	4.57	4.57	4.62	4.61
4.87	4.87	4.85	4.87	4.78
5.16	5.16	5.18	5.09	5.17
5.46	5.46	5.42	5.38	5.49
5.76	5.76	5.76	5.75	5.79
6.05	6.06	6.05	6.06	6.08
6.35	6.34	6.38	6.28	6.33
6.65	6.65	6.63	6.63	6.73
6.94	6.89	6.80	6.87	6.92
7.24	7.24	7.19	7.27	7.24
7.54	7.54	7.57	7.68	7.51
7.83	7.84	7.77	7.80	7.83
8.13	8.14	8.24	8.20	8.05
8.43	8.42	8.50	8.42	8.42
8.73	8.73	8.50	8.74	8.77
9.02	9.02	9.04	9.11	9.00
Execution time (seconds)	92.57	19.57	11.59	4.57

TABLE 2 — *125-Item, 5-Base Test Case*

Target Budget ($ millions)	Total Cost ($ millions)		
	Fox-Landi	First Approximation	Second Approximation
26.4	26.7	24.8	26.6
27.6	27.6	26.2	27.9
28.7	28.7	27.6	28.9
29.8	30.0	29.5	29.8
31.0	31.2	30.7	30.8
32.1	32.1	32.0	32.2
33.2	33.3	33.1	33.1
34.4	34.4	34.3	34.2
35.4	35.5	35.9	35.7
36.6	36.8	37.0	36.7
37.8	38.0	38.1	37.7
38.9	38.6	39.3	39.2
40.0	39.9	40.6	40.0
41.2	41.1	42.1	41.3
42.3	42.5	43.9	42.4
43.4	43.3	44.7	43.7
44.6	44.5	45.6	44.2
45.7	46.3	46.1	45.9
46.8	47.2	47.3	46.7
Execution time (seconds)	36.98	16.28	4.74

NOTE: All programs are run on an IBM 370/168.

Fox-Landi algorithm (see problems P8 and P9), the amount of computation required to solve problem P1 using these methods is directly proportional to the number of depot stock levels explicitly examined. Consequently, if this number can be reduced, then the total time required to compute an optimal solution can also be reduced.

The method of estimating the optimal depot stock level that we describe in this section is of particular value when the expected number of units in the depot resupply system for an item is 20 or more. The approximation algorithm can reduce computation time for the algorithms described in Section IV by as much as 90 percent for these high-demand items.

We have indicated how the optimal base stock level s_{ij}^* can be calculated given the depot stock level s_{i0} and the value of θ. In particular, we have shown that s_{ij}^* is optimal if it is the smallest nonnegative integer for which

$$\sum_{x<s_{ij}} p[x|\lambda_{ij}T_{ij}(s_{i0})] \leqslant \theta c_i.$$

We now develop a different but equivalent way of characterizing s_{ij}^*. To simplify notation, let us suppress the item index i. We will also assume that $p[x|\lambda_j T_j(s_0)]$ has a Poisson distribution.

Define the convex back-order function for base j as

$$B_j(s_j; s_0) \triangleq \sum_{x>s_j} (x - s_j)p[x|\lambda_j T_j(s_0)],$$

for $s_j \geqslant 0$ and integer, and $\hat{B}_j$, the piecewise linear completion of B_j, as

$$\hat{B}_j(t; s_0) \triangleq \begin{cases} B_j(t; s_0) \text{ if } t \text{ is a nonnegative integer,} \\ [B_j(s_j; s_0) - B_j(s_j - 1; s_0)][t - (s_j - 1)] \\ + B(s_j - 1; s_0),\ s_j - 1 < t < s_j; \\ \text{where } s_j \text{ is a nonnegative integer,} \\ \text{and } B(-1; s_0) \triangleq \infty. \end{cases}$$

Let

$$\Delta\hat{B}_j(s_j; s_0) \triangleq \hat{B}_j(s_j; s_0) - \hat{B}_j(s_j - 1; s_0),$$

when s_j is a nonnegative integer, and

$$D(s_j; s_0) \triangleq \{v: \Delta\hat{B}_j(s_j; s_0) < v \leqslant \Delta\hat{B}_j(s_j + 1; s_0)\}.$$

Observe that $D_j(s_j;s_0) \cup \{\Delta\hat{B}_j(s_j; s_0)\}$ is the set of subgradients of $\hat{B}_j$ at s_j. Then an alternative way of verifying that s_j^* is an optimal base stock level is to show that $-\theta c \in D(s_j^*; s_0)$.

Next let

$$F(s_1, s_2, \ldots, s_n; s_0) \triangleq \sum_{j=1}^{m} [B_j(s_j; s_0) + \theta c s_j].$$

By dropping both the integrality and nonnegativity restrictions on s_0, we obtain the following relaxation of problem P8:

(P10) $$\min_{s_0} \left\{\theta c s_0 + \min_{s_j=0,1,\ldots} \{F(s_1, \ldots, s_n; s_0): s_0 \text{ fixed}\}\right\}.$$

If s_0 is the optimal solution to problem P10, then

$$\frac{\partial F}{\partial s_0} + \theta c = 0. \tag{E3}$$

But

$$\frac{\partial F}{\partial s_0} = \sum_{j=1}^{m} \frac{\partial B_j}{\partial T_j} \frac{\partial T_j}{\partial s_0}.$$

Furthermore, by writing $B_j(s_j; s_0)$ as

$$\sum_{K=1}^{\infty} Kp[K + s_j | \lambda T_j(s_0)],$$

we see that

$$\begin{aligned}\frac{\partial B_j}{\partial T_j} &= \sum_{K=1}^{\infty} \lambda_j K e^{-\lambda_j T_j(s_0)} \frac{[\lambda_j T_j(s_0)]^{K+s_j-1}}{(K + s_j - 1)!} \\ &\quad - \sum_{K=1}^{\infty} K\lambda_j e^{-\lambda_j T_j(s_0)} \frac{[\lambda_j T_j(s_0)]^{K+s_j}}{(K + s_j)!} \\ &= -\lambda_j \Delta \hat{B}(s_j; s_0).\end{aligned}$$

As we discussed in Section II, tne function

$$B_0(s_0) \triangleq \sum_{x > s_0} (x - s_0) p(x | \lambda D)$$

can be closely approximated by an exponential function of the form $a_0 e^{-b_0 s_0}$, where a_0 and b_0 are positive real numbers. Then

$$T_j(s_0) = r_j A_j + (1 - r_j) B_j + \frac{1}{\lambda} a_0 e^{-b_0 s_0}$$

and

$$\frac{\partial T_j}{\partial s_0} = -\frac{(1 - r_j)}{\lambda} a_0 b_0 e^{-b_0 s_0}.$$

Upon combining these observations, we see that

$$\frac{\partial F}{\partial s_0} \cong \sum_{j=1}^{m} \lambda_j \Delta \hat{B}(s_j; s_0) \frac{(1 - r_j)}{\lambda} a_0 b_0 e^{-b_0 s_0}.$$

Recall that $-\theta c \in D(s_j^*; s_0)$. Consequently $-\theta c$ approximates the marginal reduction in backorders at base j when the stock level at that base is s_j^*. After making this substitution and representing this further approximation of $\partial F/\partial s_0$ by $\widehat{\partial F}/\partial s_0$, we see that

$$\begin{aligned}\frac{\widehat{\partial F}}{\partial s_0} &= -\sum_{j=1}^{m} (1 - r_j)\lambda_j \frac{1}{\lambda} \theta c a_0 b_0 e^{-b_0 s_0} \\ &= -\theta c a_0 b_0 e^{-b_0 s_0}.\end{aligned}$$

Substituting this approximation into (E3) we obtain the following estimates of the optimal depot stock level:

(E4) $$\hat{s} = -\frac{1}{b_0} \ln\left\{\frac{1}{a_0 b_0}\right\}.$$

Recall that the value of $\hat{s}_0$ is derived based on an exponential approximation of $B_0(s_0)$. As the average number of units in the depot repair cycle increases, that is, as λD increases, the quality of this exponential approximation improves in the region in which the optimal depot stock level should be located. Consequently, the approximation should be most accurate in these cases. But the problems for which the search for the optimal depot stock level is most time-consuming for the algorithms described in Section IV correspond to the items having a large number of units in depot repair. Therefore, the proposed approximation method will be most appropriate for the items requiring the greatest amount of computational effort.

The approach we have described for estimating the optimal depot stock level has been coded and tested using a sample of 40 F-15 aircraft items. The test consisted of two sets of runs. In the first set, monthly flying was divided among 3 bases; in the second set the same monthly flying program was divided among 5 bases. The total budget distributed among the 40 items ranged from $34 million to $65 million in the first set of runs, and from $34 million to $88 million in the second set. Table 3 contains the data indicating both the optimal and the estimated depot stock levels for each item in both runs.

As shown in the table, there is usually no single optimal depot stock level for an item. Rather, the optimal value depends on the amount of total item system stock available for distribution among the depot and bases. The estimate of optimal depot stock is quite close to the optimal value in all cases. Furthermore, the increase in expected system backorders using the estimated depot stock levels rather than the optimal levels is generally small. For most items, the increase is substantially less than 0.1 back orders.

The results of the tests indicate that it is possible to estimate closely the optimal depot stock level using (E4). Additionally, incorporating this method for estimating the optimal depot stock into the algorithms described in Section IV will considerably reduce the search required to find the optimal depot stock level, and will therefore markedly reduce the computational time needed to solve problem P1 using these algorithms.

REFERENCES

[1] Feeney, George J., and Craig C. Sherbrooke, "The $(s-1, s)$ Inventory Policy Under Compound Poisson Demand," Management Science, *12*, 391-411 (1966).

[2] Fox, Bennett, and M. Landi, "Searching for the Multiplier in One-Constraint Optimization Problems," Operations Research, *18*, 253-262 (1970).

[3] Sherbrooke, Craig C., "METRIC: A Multi-Echelon Technique for Recoverable Item Control," Operations Research, *16*, 122-141 (1968).

TABLE 3 — *Comparison of Optimal and Estimated Depot Stock Levels*

Item	Optimal Depot Stock Levels		Estimated Optimal Depot Stock Levels
	Case I (3 bases)	Case II (5 bases)	
1	4-7	5-9	6
2	1,2	1-3	1
3	6	6,7	6
4	0-2	2,3	1
5	10,11	8-12	10
6	18-21	18-21,25	19
7	1,2	1,2	1
8	2	3,4	2
9	5,6	6,7	6
10	1	1,2	1
11	4,5	4-6	5
12	1	1	0
13	0-2	0,1	0
14	1-3	1-3	2
15	2-4	3,4	3
16	8,9	8,9	8
17	1,2	1,2	1
18	3,4	3-5	3
19	12-14	13-14	12
20	9-12	10-13	10
21	21-27	22-28	23
22	4,5	4-6	5
23	1	1-3	1
24	1,2	2,3	2
25	5-7	6,7	6
26	16	16	16
27	3	3,4	3
28	40-42	41-43	40
29	8-10	9,10	9
30	1	2	1
31	1,2	1,2	1
32	8,9	8,9	8
33	4,5	5,6	5
34	9-11	9,10	10
35	6,7	7,8	7
36	1-3	2	2
37	1,2	1,2	1
38	7,8	7-9	7
39	2,3	3,4	3
40	41-43	42-44	41

AN INVENTORY DEPLETION PROBLEM WITH RANDOM AND AGE-DEPENDENT LIFETIMES

Daniel Thorburn

National Central Bureau of Statistics
*Stockholm, Sweden**

ABSTRACT

The following problem is studied. The units of an inventory are used one by one until all have failed. Their lifetimes decrease with their ages, when they are taken out of the inventory. An item of age a is supposed to have a lifetime $Y\exp(-a)$, where Y is a random variable which does not depend on a. It is shown that in order to maximize the total lifetime the items should be taken according to the LIFO principle. This is shown for a certain class of distributions of Y. This class includes the exponential and the Pareto distributions.

1. INTRODUCTION

Consider a stockpile consisting of n items, which are characterized by their age. When an item of age a is taken out it will last for the time $T(a)$. When it fails it is immediately replaced by a new one until there are no more items left. If, for instance, there are only two items in the stockpile, with ages a_1 and a_2, the total lifetime will be either $T(a_1, a_2) = T(a_1) + T[a_2 + T(a_1)]$ or $T(a_2, a_1) = T(a_2) + T[a_1 + T(a_2)]$, depending on the order in which the items are taken out. Lieberman [4] asked in what order the items should be taken in order to make the total field life as long as possible. He and later Brown and Ross [1] gave almost complete answers when the lifetime is a fixed function of the age.

In this paper we shall study a special case of random lifetimes. We suppose that they are distributed as

$$Y\exp(-ca), \tag{1.1}$$

where Y is a random variable and c is a constant. We will prove that the LIFO principle is optimal for a class of distributions of the random variable Y. This class includes among others the exponential distribution. LIFO is short for "Last In First Out," i.e. the youngest item in stock shall always be the one that is taken out. Ross (Ref. [5], p. 179) stated this as an open question. He conjectured that the LIFO principle would be optimal. Brown and Solomon [2] studied random lifetimes.

In the next section we introduce the notations and do some preliminary work. In the third section we discuss the paper of Brown and Solomon and its relation to our work. Our main result, Theorem 4.1, is given and proved in the following section. The last section contains two examples, in which Theorem 4.1 holds.

*Research done when the author was at the Department of Mathematical Statistics, University of Lund, Lund, Sweden.

2. PRELIMINARIES

When the lifetimes are fixed it is clear what is meant by maximal total field life. When they are random this is not so obvious. Ross [5] considered expected total field life. We will, however, use the same criterion as Brown and Solomon [2], stochastic monotonicity. A random variable X is stochastically larger (or smaller) than Y if its distribution function lies to the right of and below (or to the left of and above) that of Y. We will write this as $X \overset{d}{\geqslant} Y$ (or $X \overset{d}{\leqslant} Y$). (If they have the same distribution we use an equality sign.) It is easy to see that X is stochastically larger than Y if and only if $E[h(X)] \geqslant E[h(Y)]$ for all increasing functions h, such that the expected values exist. If we take h to be the identity, this is the criterion of Ross.

We will also use the monotone likelihood ratio (m.l.r.). The random variable X has increasing (or decreasing) likelihood ratio with respect to Y if $f_x(t)/f_Y(t)$ increases (or decreases) with t. Here f_X and f_Y denote the densities of X and Y. It is easy to see that the monotone likelihood ratio implies stochastic monotonicity. If every pair of random variables in a family has a monotone likelihood ratio, we will say that the family has the m.l.r. property. An equivalent definition of m.l.r. is $f_X(x)\, f_Y(y) - f_X(y)\, f_Y(x) \geqslant 0$ whenever $x \geqslant y$.

We will also use failure rates [3]. Suppose that the random variable X has the distribution function F and the density f. Its failure rate is

$$\mu(t) = f(t)/[1 - F(t)] = \lim_{h \to 0} P(X \leqslant t+h \mid X > t)/h.$$

An absolutely continuous distribution is completely defined by its failure rate:

$$F(x) = 1 - \exp[-\int_{\infty}^{x} \mu(t)\, dt].$$

In the following we shall only study positive random variables. If a distribution is defined by a failure rate such that $F(x)$ does not converge to one, we say that the random variable takes the value infinity with probability $\exp[-\int \mu(t)\, dt]$. Some results for failure rates, which are needed later on, are given below.

LEMMA 2.1: Let X and Y be two independent random variables with failure rates μ_X and μ_Y. Then the failure rate μ of $\min(X, Y)$ is

$$\mu(t) = \mu_X(t) + \mu_Y(t).$$

PROOF: The failure rate $\mu(t)$ equals

$$\lim_{h \to 0} \frac{1}{h} P(X \leqslant t+h \text{ or } Y \leqslant t+h \mid X > t \text{ and } Y > t)$$

$$= \lim_{h \to 0} \frac{1}{h} [P(X \leqslant t+h \mid X > t) + P(Y \leqslant t+h \mid Y > t)$$

$$- P(X \leqslant t+h \mid X > t) \cdot P(Y \leqslant t+h \mid Y > t)]$$

$$= \mu_X(t) + \mu_Y(t) - 0.$$

LEMMA 2.2: If X and Y are two positive random variables such that X has increasing likelihood ratio with respect to Y then $\mu_X(x) \leqslant \mu_Y(x)$ for all positive x.

PROOF: The proof is by contradiction. Suppose that for some value x_o,

$$\mu_X(x_o) = \frac{f_X(x_o)}{\int_{x_o}^{\infty} f_X(t)\,dt} > \frac{f_Y(x_o)}{\int_{x_o}^{\infty} f_Y(t)\,dt} = \mu_Y(x_o).$$

Using the assumption of m.l.r., we have the following contradiction,

$$1 = \int_{x_o}^{\infty} \left\{ \left[\frac{f_X(x)}{\int_{x_o}^{\infty} f_X(t)\,dt} \right] \Big/ \left[\frac{f_Y(x)}{\int_{x_o}^{\infty} f_Y(t)\,dt} \right] \right\} \left[\frac{f_Y(x)}{\int_{x_o}^{\infty} f_Y(t)\,dt} \right] dx$$

$$> \int_{x_o}^{\infty} \frac{f_Y(x)}{\int_{x_o}^{\infty} f_Y(t)\,dt}\,dx = 1.$$

If instead of two random variables we consider a multiplicative family with the m.l.r. property we get the following corollary.

COROLLARY 2.1: If X is a random variable and if the family aX has the m.l.r. property then $x\mu_X(x)$ increases with x.

PROOF: The failure rate of aX is $\mu_X(x/a)/a$. By the previous lemma this is decreasing in a and the result follows.

LEMMA 2.3: Let X be a random variable. Let $h(\cdot)$ be a convex function such that $h(0) = 0$ and $h(x) - x$ is increasing. If $x\mu_X(x)$ increases with x then $\mu_X(x) \geq \mu_{h(X)}(x)$.

PROOF: Since $h(x)$ is convex it must be differentiable almost everywhere and $h(t) = \int_0^t h'(x)\,dx$. Define $h'(t)$ to be right continuous in the points where the derivative does not exist. The function $h'(x)$ so defined is increasing in x because $h(x)$ is convex. It is now easy to see that

$$h(t)/t = \int_0^t h'(x)\,dx/t \leq t \cdot h'(t)/t = h'(t).$$

The conditions of the lemma give that $h(t) \geq t$ and hence

$$t\mu_X(t) \leq h(t)\,\mu_X[h(t)].$$

Using these two formulas we have

$$\mu_{h(X)}[h(t)] = \mu_X(t)/h'(t) \leq \mu_X[h(t)]\,h(t)/[t\,h'(t)]$$
$$\leq \mu_X[h(t)].$$

With $x = h(t)$ the result follows.

3. RELATION TO PREVIOUS WORK

Brown and Solomon [2] studied random lifetimes, but they did not consider exponential deterioration for other than bounded variables. Let $X_1 \overset{d}{\geq} X_2 \overset{d}{\geq} \cdots \overset{d}{\geq} X_n$ be independent random variables and d a positive function. An item i will last for the time $X_i d(t)$ if it is kept in stock for the time t before it is taken out.

Brown and Solomon showed that if d is increasing and convex the items should be taken in the given order. If on the other hand d is increasing and concave they should be taken in the opposite order. They also showed that if d is convex and $P(X_1 \geq X_2) = 1$, the items should be taken out with increasing indexes. We will here give a simple corollary of this result. The corollary says that if all the lifetimes are bounded by M and if $d(t)$ does not decrease faster than $1/M$, then the items should be taken with increasing indexes.

THEOREM 3.1: Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables such that $X_1 \overset{d}{\geq} X_2 \overset{d}{\geq} \ldots \overset{d}{\geq} X_n$ and $X_2 < M$ with probability one. If d is positive decreasing and convex and if $|d'(0)| < 1/M$ then the items should be taken in the given order.

PROOF: If $X_1 \geq M$, the result follows immediately from the results of Brown and Solomon, so let us assume that $X_1 < M$. Let $X_0 = M$ and define $g(t)$ by

$$\begin{cases} \dfrac{d(0)}{1 + Md'(0)} + d'(0)t & \text{if } t \leq \dfrac{Md(0)}{1+Md'(0)} \\ d\left[t - \dfrac{Md(0)}{1 + Md'(0)}\right] & \text{if } t \geq \dfrac{Md(0)}{1 + Md'(0)} \end{cases}$$

The result of Brown and Solomon can now be applied to $X_0, \ldots, X_n$ and $g(t)$. Thus $X_0, X_1, \ldots, X_n$ should in this situation be taken in this order. But when X_0 fails we have exactly the situation described by our theorem and the result follows.

Brown and Solomon also proved that if there are only two items with m.l.r. in the stockpile, the larger one should be taken first when d is positive and convex. We will here prove a weaker result using a proof that illustrates one of the tricks used in proving Theorem 4.1. First we state a simple lemma without proof.

LEMMA 3.1: The function $x_1 + x_2 \exp(-cx_1) - x_2 - x_1 \exp(-cx_2)$ is positive, increasing, and convex as a function of $x_1 \geq x_2$.

Denote the initial ages of the items in stock by $a_1 \leq a_2$ and the total field life by $T(a_1, a_2)$, if the item with initial age a_1 is taken first, and by $T(a_2, a_1)$ if the items are taken in the opposite order. We consider from now on the situation described by (1.1), i.e., with exponential deterioration.

LEMMA 3.2: Let the family aY, $a \in R^+$, have the m.l.r. property. Then $T(a_1, a_2) \overset{d}{\geq} T(a_2, a_1)$.

PROOF: Let X_1 and X_2 be independent random variables such that $X_1 \overset{d}{=} Y \exp(-ca_1)$ and $X_2 \overset{d}{=} Y \exp(-ca_2)$. With this definition we have

$$\text{(3.1)} \qquad \begin{cases} X_1 + X_2 \exp(-cX_1) \overset{d}{=} T(a_1, a_2) \\ X_2 + X_1 \exp(-cX_2) \overset{d}{=} T(a_2, a_1) \end{cases}$$

If the densities of X_1 and X_2 are denoted by f_1 and f_2, respectively, then their joint density can be written

$$\text{(3.2)} \qquad \begin{aligned} & f_1(x_1)\, f_2(x_2) \\ &= [f_1(x_1)\, f_2(x_2)\, I(x_1 \leq x_2) + f_1(x_2)\, f_2(x_1)\, I(x_1 > x_2)] \\ &\quad + [f_1(x_1)\, f_2(x_2) - f_1(x_2)\, f_2(x_1)]\, I(x_1 > x_2). \end{aligned}$$

where I is the indicator function. The first term is symmetric in x_1 and x_2. The second term is positive when $x_1 > x_2$ and vanishes elsewhere. This follows from the assumption of m.l.r.

In order to obtain the distributions of $T(a_1, a_2)$ and $T(a_2, a_1)$, we integrate the two terms over the sets $x_1 + x_2 \exp(-cx_1) \leqslant t$ and $x_2 + x_1 \exp(-cx_2) \leqslant t$, respectively. From the symmetry it follows that the integrals of the first term are equal. It follows from Lemma 3.1 that the integral of the second term is smaller over the first set. Lemma 3.2 now follows from the definition of stochastic monotonicity.

REMARK 3.1: This method of dividing the density functions can also be used when $n = 3$. In that case, however, the density has to be divided into six rather complicated terms, all of which require different proofs. We will not try to do so here since, as far as we know, there is no possibility of generalizing this method to a general n.

4. A GENERAL NUMBER OF ITEMS

Let the items in stock have ages $a_1 \leqslant a_2 \leqslant a_3 \leqslant \ldots \leqslant a_n$ and let $T(a_{i(1)}, a_{i(2)}, \ldots, a_{i(n)})$ denote the total field life when the item with age $a_{i(1)}$ is taken first, that with initial age $a_{i(2)}$ is taken next, and so on. If we assume that the LIFO principle is optimal when the stockpile consists of p items, and if $n = p + 1$ we know that the p last items will be taken according to the LIFO principle. The youngest one will thus be one of the first two items taken. If we could show that

$$\begin{aligned} &T(a_1, a_k, a_2, \ldots, a_{k-1}, a_{k+1}, \ldots, a_{p+1}) \\ &\overset{d}{\geqslant} T(a_k, a_1, a_2, \ldots, a_{k-1}, a_{k+1}, \ldots, a_{p+1}) \end{aligned} \tag{4.1}$$

the result would follow, for by optimality of the LIFO principle when $n = p$, the first quantity is always stochastically smaller than $T(a_1, a_2, \ldots, a_{p+1})$.

Formula (4.1) can be rewritten as

$$\begin{aligned} &T(a_1, a_k) + T[a_2 + T(a_1, a_k), \ldots, a_{k-1} + T(a_1, a_k), \\ &\qquad a_{k+1} + T(a_1, a_k), \ldots, a_{p+1} + T(a_1, a_k)] \\ &\overset{d}{\geqslant} T(a_k, a_1) + T[a_2 + T(a_k, a_1), \ldots, a_{k-1} + T(a_k, a_1), \\ &\qquad a_{k+1} + T(a_k, a_1), \ldots, a_{p+1} + T(a_k, a_1)]. \end{aligned} \tag{4.2}$$

The problem can thus be viewed as a p-unit problem, where the first item has a different distribution (here: $T(a_1, a_k)$ or $T(a_k, a_1)$, respectively). Lemma 4.1 treats this problem. We repeat that μ is the failure rate of the variable Y in the definition (1.1) of lifetime distributions.

LEMMA 4.1: Suppose that the LIFO principle is optimal when $n = p$. Let U_1 and U_2 be random variables with failure rates μ_1 and μ_2 and let the following conditions hold:

$$\mu\left\{x \exp[c(y + a_1)]\right\} \geqslant \mu[(x + y) \exp(cy)] \exp(ca_1), \tag{4.3}$$

$$\mu_1(x) \leqslant \mu_2(x), \text{ and} \tag{4.4}$$

$$\mu_1(x) \leqslant \mu[x \exp(ca_1)] \exp(ca_1). \tag{4.5}$$

The following stochastic inequality holds:

$$T_1 = U_1 + T(a_1 + U_1, \ldots, a_{p-1} + U_1)$$

$$\overset{d}{\geqslant} T_2 = U_2 + T(a_1 + U_2, \ldots, a_{p-1} + U_2).$$

PROOF: Let us suppose that we have the situation described as T_1, i.e. p units such that the first unit has the lifetime U_1 and the other units have lifetime distributions corresponding to their initial ages $a_1, a_2, \ldots, a_{p-1}$.

Construct Z_1 with the failure rate $\lambda_1(t) = \mu_2(t) - \mu_1(t)$, which by (4.4) is positive. Now by Lemma 2.1 we have $U_2 \overset{d}{=} \min(Z_1, U_1)$. If we replace the first unit at time $\min(Z_1, U_1)$, even though it may be functioning, the total field life will be distributed as T_2. The general idea of the proof is to change $\min(Z_1, U_1)$ into U_1 in a number of steps so that each step will yield a stochastically larger field life.

If $Z_1 \geqslant U_1$, we are finished. If $Z_1 < U_1$, we define the variable U_3 as

$$U_3 = U_1 - Z_1. \tag{4.6}$$

The failure rate of U_3 is $\mu_1(t + Z_1)$. If $Z_1 < U_1$ the remaining total field life after Z_1 is distributed as

$$T(a_1 + Z_1, \ldots, a_{p-1} + Z_1) \tag{4.7}$$

or as

$$U_3 + T(a_1 + Z_1 + U_3, \ldots, a_{p-1} + Z_1 + U_3) \tag{4.8}$$

depending on whether we replace the first item at Z_1 or not.

Construct a new random variable Z_2 depending on Z_1 by its failure rate

$$\lambda_2(t) = \mu\left\{t \exp[c(a_1 + Z_1)]\right\} \exp[c(a_1 + Z_1)] - \mu_1(t + Z_1), \tag{4.9}$$

which by conditions (4.3) and (4.5) is positive. This definition leads by Lemma 2.1 to

$$\min(U_3, Z_2) \overset{d}{=} Yg(a_1 + Z_1). \tag{4.10}$$

If we replace neither at Z_1 as in (4.7) nor at $U_1 = Z_1 + U_3$ as in (4.8) but at $Z_1 + \min(U_3, Z_2)$, the remaining total field life at Z_1 will be distributed as

$$T(a_1 + Z_1, a_1 + Z_1, a_2 + Z_1, \ldots, a_{p-1} + Z_1).$$

By the induction hypothesis this is stochastically larger than

$$T(a_1 + Z_1, \ldots, a_{p-1} + Z_1, a_1 + Z_1) \geqslant T(a_1 + Z_1, \ldots, a_{p-1} + Z_1).$$

In other words, we have shown that the total field life is stochastically longer if the first unit has the lifetime

$$\min(Z_1, U_1) + \min(Z_2, U_3)\, I(Z_1 < U_1) = \min(Z_1 + Z_2, U_1), \tag{4.11}$$

than if the first unit has the lifetime

$$\min(Z_1, U_1) \overset{d}{=} U_2. \tag{4.12}$$

The first step in the change of $\min(Z_1, U_1)$ into U_1 is now completed. The next step will proceed in exactly the same way with Z_1 replaced by $Z_1 + Z_2$. If $Z_1 + Z_2 \geqslant U_1$, we are

finished; otherwise, we define $U_4 = U_1 - Z_1 - Z_2$ [see (4.6)]. It has the failure rate $\mu_1(t + Z_1 + Z_2)$. Construct Z_3 depending on Z_1 and on Z_2 with the failure rate [see (4.9)]

$$\lambda_3(t) = \mu\left\{t \exp\left[c(a_1 + Z_1 + Z_2)\right]\right\} \exp\left[c(a_1 + Z_1 + Z_2)\right] - \mu_1(t + Z_1 + Z_2).$$

Now [see (4.10)] we have $\min(U_4, Z_3) \overset{d}{=} Y\, g(a_1 + Z_1 + Z_2)$. We thus find that the total field life is stochastically larger if the first item has the lifetime [see (4.11) and (4.12)] $\min(Z_1 + Z_2 + Z_3, U_1)$ than if it has the lifetime U_2.

The second step is completed. By repeating this argument unitl $Z_1 + Z_2 + \ldots \geqslant U_1$ we find that it would be better to let the first item remain in the system until it fails, i.e. until the time U_1.

That $Z_1 + Z_2 + \ldots$ eventually will reach U_1 follows from the fact that $\lambda_i(t)$ is bounded in every finite interval

$$\lambda_i(t) \leqslant \sup_{0 \leqslant x \leqslant t} \mu[t \exp(cx)] \exp(cx) \quad \text{for} \quad i > 1.$$

The distribution of U_2 is thus changed into that of U_1 in such a way that the total field life increases stochastically in each step.

THEOREM 4.1: If the following conditions hold:

(4.13) $\mu(x)$ decreases in x,

(4.14) $Y \cdot a$, $a \in R^+$ has the m.l.r. property, and

(4.15) $x\,\mu'(x)/\mu(x)$ decreases in x,

the LIFO principle is optimal.

PROOF: It is trivial that the theorem holds for $n = 1$. We have in fact, by Lemma 3.2, that it holds for $n = 2$, but we do not need that fact. We will now prove that if it holds for $n = p$, then it also holds for $n = p + 1$. When this is done the result follows from the axiom of induction. Unfortunately it is not possible to use Lemma 4.1 directly on (4.2). We will therefore first change the situation into one where the lemma is applicable and then check the conditions. Exactly as in the proof of Lemma 3.2 we find that it is sufficient to consider the situation

$$\begin{cases} X_1 + X_k \exp(-cX_1) \overset{d}{=} T(a_1, a_k) \\ X_k + X_1 \exp(-cX_k) \overset{d}{=} T(a_k, a_1) \end{cases}$$

where the distribution of X_1 and X_k is given by $K[f_1(x_1)\, f_k(x_k) - f_1(x_k)\, f_k(x_1)]\, I(x_1 > x_k)$. In order to simplify notation we introduce the following positive variables:

$$Z = X_k + X_k \exp(-cX_k),$$
$$Y_1 = X_1 + X_k(-cX_1) - Z, \text{ and}$$
$$Y_2 = (X_1 - X_k) \exp(-cX_k).$$

Formula (4.2) is now reduced to

$$Y_1 + T(a_2 + Z + Y_1, \ldots, a_{k-1} + Z + Y_1, a_{k+1} + Z + Y_1, \ldots, A_{p+1} + Z + Y_1)$$
$$\overset{d}{\geqslant} Y_2 + T(a_2 + Z + Y_2, \ldots, a_{k-1} + Z + Y_2, a_{k+1} + Z + Y_2, \ldots, a_{p+1} + Z + Y_1).$$

This formula will be shown to hold for any given fixed Z by use of Lemma 4.1. Observe that a_i in the lemma corresponds to $a_{i+1} + Z$ or $a_{i+2} + Z$ here, depending on whether $i \leqslant k - 1$ or not. The failure rate of Y_1 and Y_2, given X_k, are called μ_1 and μ_2.

Condition (4.3) is easily shown to follow from (4.13) and (4.14) and Corollary 2.1:

$$\exp\left[c(a_2 + Z + y)\right] \mu\left\{x \exp\left[c(a_2 + Z + y)\right]\right\}$$
$$\geqslant \exp\left[c(a_2 + Z + y)\right] \mu\left\{(x + y) \exp\left[c(a_2 + Z + y)\right]\right\}$$
$$\geqslant \exp\left[c(a_2 + Z)\right] \mu\left\{(x + y) \exp\left[c(a_2 + Z)\right]\right\}.$$

Condition (4.4) is more complicated. If $x\mu_2(x)$ were increasing, Lemma 3.1 would say that $Y_1 - Y_2$ is positive, increasing, and convex as a function of Y_2, and Lemma 2.3 that condition (4.4) held for such functions. We will thus try to show that

$$(x - X_k) \frac{f_1(x) f_k(X_k) - f_1(X_k) f_k(x)}{\int_x^\infty f_1(t) f_k(X_k) - f_1(X_k) f_k(t)\, dt} =$$

(4.16)

$$\left\{\mu\left[(x - X_k) \exp(ca_1)\right] x \exp(ca_1)\right\} \frac{1 - \dfrac{f_k(x)}{f_1(x)} \dfrac{f_1(X_k)}{f_K(X_k)}}{1 - \dfrac{\mu[x \exp(ca_1)]}{\mu[x \exp(ca_k)]} \dfrac{\exp(ca_1)}{\exp(ca_k)} \dfrac{f_k(x)}{f_1(x)} \dfrac{f_1(X_k)}{f_k(X_k)}}.$$

is increasing.

By Corollary 2.1 the first factor is increasing. The m.l.r. property gives that the second term of the numerator decreases from one as x increases from X_k. That

$$\frac{\mu[x \exp(c\ a_1)] \exp(c\ a_1)}{\mu[x \exp(c\ a_k)] \exp(c\ a_k)} \tag{4.17}$$

is less than one and increasing in x follows from Lemma 2.2 and the fact that the derivative of its logarithm

$$\exp(ca_1) \frac{\mu'[x \exp(ca_1)]}{\mu[x \exp(ca_1)]} - \exp(ca_k) \frac{\mu'[x \exp(ca_k)]}{\mu[x \exp(ca_k)]},$$

is positive [by (4.15)].

That the second factor of (4.16) is also increasing can now be seen by writing it as

$$(1 - y)/[1 - h(y)\, y], \tag{4.18}$$

where y equals the second term of the numerator of (4.16) and $h(y)$ equals (4.17). Formula (4.18) is easily seen to be increasing in x noting that its derivative with respect to y is negative. We have now shown that (4.16) is increasing in $x > X_k$.

Condition (4.5) remains to be proved. If $x > X_k$ the failure rate of X_1 equals

$$\mu_2\Big[(x - X_k)\exp(-cX_k)\Big]\exp(-cX_k)$$

$$= \frac{f_1(x)\,f_k(x_k) - f_1(x_k)\,f_k(x)}{\int_x^\infty f_1(t)\,dt\,f_k(x_k) - f_1(x_k)\int_x^\infty f_k(t)\,dt}$$

$$\leqslant f_1(x)/\int_x^\infty f_1(t)\,dt$$

$$= \mu[x\exp(ca_1)]\exp(ca_1).$$

The first expression follows from the definition of Y_2. The inequality is a consequence of the m.l.r. property and Lemma 2.2. A redefinition of x and (4.13) and (4.14) gives

$$\mu_1(x) \leqslant \mu_2(x)$$

$$\leqslant \mu\Big\{x\exp\Big[c(a_1 + X_k)\Big] + X_k\exp(ca_1)\Big\}\exp\Big[c(a_1 + X_k)\Big]$$

$$\leqslant \mu\Big\{x\exp\Big[c(a_2 + X_k)\Big]\Big\}\exp\Big[c(a_2 + X_k)\Big]$$

Condition (4.5) follows if we remember that a_1 in Lemma 4.1 corresponds to $a_2 + X_k$ in this proof.

5. APPLICATIONS

It is easy to check that the conditions of Theorem 4.1 hold for the exponential distribution. The failure rate is in that case a constant. Condition (4.13) that $\mu(x)$ is decreasing and condition (4.14) that $x\,\mu'(x)/\mu(x)$ is decreasing are both trivial. Condition (4.15) that the exponential distribution has the m.l.r. property is well known.

Another class of distributions for which the conditions hold is the Pareto family starting at zero:

$$f(x) \propto (x + a)^{-k},\ a > 0,\ k > 1. \tag{5.1}$$

The failure rate of this distribution equals

$$\mu(x) = (k - 1)/(x + a). \tag{5.2}$$

It is easily seen that (5.2) is decreasing in x. The m.l.r. property holds since

$$f_\gamma(x)/f_{a\gamma}(x) = (x - 1)^{-k}/[(1/a)\,(x/a - 1)]^{-k}$$

$$= [1 + (a - 1)/(x - a)]^{-k}\,a^{1-k}$$

is increasing in x if $a \geqslant 1$. Condition (4.15) is easy to check:

$$x\mu'(x)/\mu(x) = x[-(k - 1)/(x + a)^2]/[(k - 1)/(x + a)]$$

$$= -1 + a/(x + a).$$

REFERENCES

[1] Brown, M., and Ross, S. M., "Optimal Issuing Policies," Management Science, *19*, 1292-1294 (1973).

[2] Brown, M., and Solomon, H., "Optimal Issuing Policies Under Stochastic Field Lives," Journal of Applied Probability, *10*, 761-768 (1973).
[3] Cox, D. R., *Renewal Theory*, (Methuen, London, 1962).
[4] Lieberman, G. J., "LIFO vs FIFO in Inventory Depletion Management," Management Science *5*, 102-105 (1958).
[5] Ross, S. M., *Applied Probability Models with Optimization Applications* (Holden-Day, San Francisco, 1969).

APPROXIMATING PARTIAL INVERSE MOMENTS FOR CERTAIN NORMAL VARIATES WITH AN APPLICATION TO DECAYING INVENTORIES*

Steven Nahmias and Shan Shan Wang

University of Pittsburgh
Pittsburgh, Pennsylvania

ABSTRACT

This paper considers the problem of computing $E(X^{-n}; X > t)$ when X is a normal variate having the property that the mean is substantially larger than the standard deviation. An approximation is developed which is determined from the mean, standard deviation, and the cumulative standard normal distribution. Computations comparing the approximate moments with the actual are reported for various values of the relevant parameters. These results are applied to the problem of computing the expected number of shortages in a lead-time for a single product which exhibits continuous exponential decay.

1. THE APPROXIMATION

The purpose of this paper is to consider an approximation for $E(X^{-n}; X > t)$ for $t > 0$ when X is a normal variate with the property that $\mu >> \sigma$, which is common in physical processes in which a normal distribution is used to describe a nonnegative phenomenon. Since the probability of obtaining a negative observation is $\Phi(-\mu/\sigma)$ (where Φ is the cumulative normal distribution) it must be true that $\mu >> \sigma$ in order that this probability be neligible. The application of our results to a problem associated with a product which exhibits continuous exponential decay will also be considered. The random variable X will correspond to the lead-time demand which is a nonnegative random variable generally assumed to be normal.

We have that

$$E(X^{-n}; X > t) = \int_t^\infty x^{-n} f(x \mid \mu, \sigma)\, dx \tag{1}$$

where

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -(x-\mu)^2/2\sigma^2 \right\}.$$

For any $n \geq 1$, this computation would require using numerical integration methods. The approximation developed here is based on approximating the normal density by a gamma density, simplifying, and then reapproximating the resulting gamma density by the normal.

Since $\mu >> \sigma$, $f(x)$ can be approximated by

*Research Supported by the National Science Foundation under grant ENG 75-04990.

$$g(x \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where

$$\alpha = (\mu/\sigma)^2 \text{ and } \beta = \mu/\sigma^2.$$

The justification for this is as follows. It is well known that $g(x \mid \alpha, \beta)$ is the probability density of the sum of k independent and identically distributed random variables with denstiy $g(x \mid \alpha/k, \beta)$ (see DeGroot [1], p. 237). It follows from the central limit theorem that for α large, $g(x \mid \alpha, \beta)$ can be approximated by a normal density. The mean and variance of a Gamma random variable with parameters α and β are $\mu = \alpha/\beta$ and $\sigma^2 = \alpha/\beta^2$. Solving for α and β in terms of μ and σ^2 gives the results above. Note that the original assumption that $\mu >> \sigma$ implies that α will be large and $g(x \mid \alpha, \beta)$ will give a good approximation to $f(x \mid \mu, \sigma)$.

Hence

$$\begin{aligned} E(X^{-n}; X > t) &\approx \int_t^{\infty} x^{-n} g(x \mid \alpha, \beta)\, dx \\ &= \int_t^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1-n} e^{-\beta x} dx \\ &= \frac{\beta^n}{(\alpha-1) \ldots (\alpha-n)} \int_t^{\infty} \frac{\beta^{\alpha-n}}{\Gamma(\alpha-n)} x^{\alpha-n-1} e^{-\beta x}\, dx. \end{aligned} \tag{2}$$

The integrand of (2) is simply the gamma density $g(x \mid \alpha - n, \beta)$. Since $\mu >> \sigma$, this gives $\alpha >> 0$ which would imply that $\alpha - n >> 0$ for moderate values of n (that is, $\mu >> \sigma\sqrt{n}$). Hence, we may than reapproximate $g(x \mid \alpha - n, \beta)$ by a normal density with mean $\hat{\mu} = \frac{\alpha-n}{\beta}$ and variance $\hat{\sigma}^2 = \frac{\alpha-n}{\beta^2}$. It follows that

$$\begin{aligned} E(X^{-n}, X > t) &\approx \frac{\beta^n}{(\alpha-1) \ldots (\alpha-n)} \int_t^{\infty} f(x \mid \hat{\mu}, \hat{\sigma})\, dx \\ &= \frac{\beta^n}{(\alpha-1) \ldots (\alpha-n)} \cdot \left\{ \bar{\Phi} \left[\frac{t-\hat{\mu}}{\hat{\sigma}} \right] \right\} \\ &= \frac{\mu^n}{\prod_{i=1}^{n} (\mu^2 - i\sigma^2)} \cdot \left[\bar{\Phi} \left\{ \frac{\mu t - \mu^2 + n\sigma^2}{\sigma \sqrt{\mu^2 - n\sigma^2}} \right\} \right] \end{aligned} \tag{3}$$

where $\bar{\Phi}(x) = 1 - \Phi(x)$.

The value of n should be small enough so that $\mu > \sigma\sqrt{n}$. Otherwise the argument of $\bar{\Phi}$ in (3) will not be defined. This is only a mild restriction since (1) will be essentially zero for large n. The quality of the approximation will clearly depend upon whether or not $\alpha - n$ is sufficiently large to apply the central limit theorem and reapproximate $g(x \mid \alpha, \beta)$ with $f(x \mid \hat{\mu}, \hat{\sigma})$.

In order to compare the effectiveness of (3) as an approximation to (1) we have estimated (1) by numerical integration. Table 1 presents a summary of some of the computational results for various values of μ, σ, t, and n. Note that for $\mu/\sigma = 10$, the agreement is extremely close, while the quality of the approximation deteriorates as μ/σ decreases. It should be noted that for larger values of n, (2) may be used directly if tables of the cumulative gamma distribu-

TABLE 1 — *Comparison of Exact and Approximate Calculations*

μ	σ	t	n	Exact value (1)	Approximation (3)
10	1	1	1	0.10103158	0.10101009
			5	1.175×10^{-5}	1.165×10^{-5}
		10	1	0.04664270	0.04648207
			5	3.57×10^{-6}	3.55×10^{-6}
10	2	1	1	0.10461799	0.10416607
			5	2.931×10^{-5}	1.915×10^{-5}
		10	1	0.04375230	0.04382718
			5	2.75×10^{-6}	2.56×10^{-6}
10	3	1	1	0.11252050	0.10963436
			5	6.0250×10^{-4}	5.102×10^{-5}
		10	1	0.04133040	0.04156939
			5	2.23×10^{-6}	1.13×10^{-6}
10	4	1	1	0.12173662	0.11646283
			5	2.61×10^{-3}	3.31×10^{-4}
		10	1	0.03925619	0.0397417
			5	1.87×10^{-6}	4.89×10^{-8}

tion are available. However, in most cases where μ is relatively large, it is generally true that $\sigma < \sqrt{\mu}$ (with equality holding when X is approximately Poisson), so that the approximation should give excellent results.

2. DECAYING INVENTORIES

When inventory levels decline by a fixed fraction each period due to spoilage or loss in value (exclusive of demand), the inventory exhibits continuous exponential decay. The term exponential decay arises in the following manner. Suppose $I(t)$ represents the inventory level for a continuous review system at time t. Ignoring demand, the decay assumption may be expressed as $I(t+s)/I(t) = \gamma^s$ where γ = the fraction of stock decaying per unit time $(0 < \gamma \leqslant 1)$ and $t,s > 0$. This is equivalent to

$$I(t+s) = \exp(-\theta s)\, I(t) \tag{4}$$

where $\theta = -\ln(\gamma)$. The constant θ is ordinarily referred to as the decay rate and represents the instantaneous rate of decrease of the relative inventory level $I(t+s)/I(t)$ at $s = 0$ (that is, $\frac{d}{ds}\left[\frac{I(t+s)}{I(t)}\right]\Big|_{s=0} = -\theta$).

Decaying inventories arise in various situations. One example is radioactive materials, such as nuclear medicines (see Refs. [2] and [6]) or UO_2 nuclear fuel for a fission reactor. Another possible application for an inventory model with decay is in the area of cash flow management. Continuous discounting of future returns and continuous exponential decay correspond to precisely the same physical process. Alternatively, a decay model might be useful for providing an approximation to the more complex problem of managing a fixed-life per-

ishable commodity (see Nahmias [5], for example, for an analysis of the periodic review version of this problem).

There is an alternative interpretation of the decay model which explains more clearly how it relates to a perishable inventory problem. Suppose $I(0) = N$ units are on hand initially at time $t = 0$, and also assume that the lifetime of each of the units is a random variable having the negative exponential distribution. That is, if $T_1, T_2, \ldots, T_N$ are the respective lifetimes of each of then N items, the $P\{T_i > s\} = \exp(-\theta s)$ for $s \geqslant 0$ and $1 \leqslant i \leqslant N$. It follows that the number of units on hand at any time $s > 0$ is a random variable having the binomial distribution with parameters N and $p = \exp(-\theta s)$, so that the expected inventory level at time s, say $I(s)$, is given by Np or $I(0) \cdot \exp(-\theta s)$ which agrees with (4) for $t = 0$. It is important to recognize that the exponential decay process is a purely deterministic process which arises as the expected value of a process with random lifetimes. For large N, the law of large numbers and the concept of ensemble averages provides a justifiction for using the deterministic exponential decay model for a problem of this type.

A common inventory policy for continuous review inventory systems is to trigger an order when the level of on-hand inventory reaches r, the reorder point (see Hadley and Whitin [4], Chapter 4). An important part of determining the proper value of r is to compute the expected number of units which go short during the leadtime, say τ. Under the common assumption that leadtime demand is a normal random variable, we will show how the results of the previous section can be used to give a very close approximation for the expected shortage in leadtime for a decaying inventory.

Ghare and Schrader [3] have analyzed the extension of the simple EOQ model to allow for continuous exponential decay. They show that if the demand rate at time u is given by the function $D(u)$, $0 \leqslant u \leqslant t$, then $I(t) = e^{-\theta t} \cdot (I(0) - \int_0^t D(u)\, e^{\theta u}\, du)$. When $D(u) = \lambda$, independent of u, this reduces to

$$I(t) = e^{-\theta t} \cdot (I(0) + \lambda/\theta) - \lambda/\theta. \tag{5}$$

Suppose that X = demand in a leadtime τ. Then the demand rate during leadtime is a random variable $\lambda(X) = X/\tau$ which now must be treated as an explicit function of the leadtime demand X. We will assume that items leave stock at a constant rate of $\lambda(X)$ during leadtime. If T^* is the amount of time that elapses from the time an order is placed until inventory level hits zero, then for any realization X of the leadtime demand, the inventory level $I(t)$ will decrease according to (5) with $\lambda = \lambda(x)$, on the interval between the time the order is placed until min (T^*, τ). The case $T^* < \tau$ is pictured in Fig. 1.

It follows from (5) that for a given value of x, T^* solves

$$0 = e^{-\theta T^*}(r + \lambda(x)/\theta) - \lambda(x)/\theta$$

or

$$T^* = (1/\theta) \cdot \ln(1 + r\theta/\lambda(x)). \tag{6}$$

In order for a shortage condition to occur it is necessary that $T^* \leqslant \tau$. Let $w(r)$ be the minimum number of units demanded in a leadtime to ensure that a shortage condition will occur. Then $w(r)$ may be obtained from (6) with $T^* = \tau$ and $\lambda(x) = w(r)/\tau$, which gives

$$\tau = \frac{1}{\theta} \ln(1 + r\,\theta\,\tau/w(r))$$

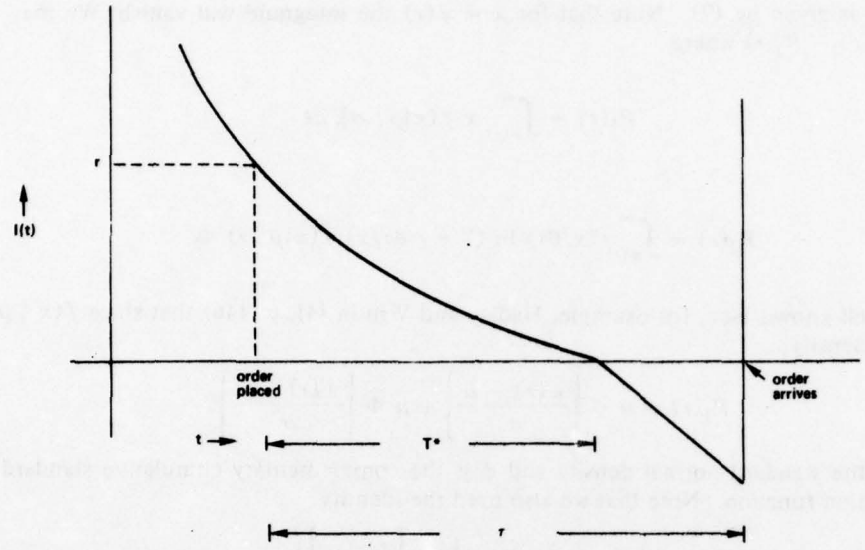


FIGURE 1 — A possible realization of the inventory level, $I(t)$, in a leadtime.

or

$$w(r) = r\,\theta\,\tau/(e^{\theta\tau} - 1). \tag{7}$$

Hence, if $X \leqslant w(r)$, then $T^* \leqslant \tau$ and no shortage occurs, while if $X > w(r)$, then $T^* < \tau$ and a shortage will occur.

Suppose that $y(x)$ = numbers of units which decay in a leadtime when x is the leadtime demand. When $T^* < \tau$, the inventory level drops from r to 0 in a time T^*, which includes the losses due to both demand and decay. The loss due to demand is exactly $\lambda(x) \cdot T^*$. It follows that the loss due to decay, $y(x)$, is given by $y(x) = r - \lambda(x)\,T^*$, which from (6) becomes

$$y(x) = r - \frac{\lambda(x)}{\theta} \ln\,(1 + r\theta/\lambda(x)). \tag{8}$$

It now follows that for all $x \geqslant w(r)$, the number of units which go short in a leadtime, say $S(x)$, is

$$S(x) = x + y(x) - r,$$

which becomes, after we substitute $\lambda(x) = x/\tau$, and use (8),

$$S(x) = x - (x/\theta\tau) \cdot \ln\,(1 + r\theta\tau/x). \tag{9}$$

Since total demand in leadtime is assumed to be normal with mean μ and variance σ^2, we obtain the expected number of units short in the leadtime, say $P(r)$, from (9) as

$$P(r) = E(S(X),\ X \geqslant w(r)) =$$

$$\int_{w(r)}^{\infty} \left\{x - (x/\theta\tau) \cdot \ln\,(1 + r\,\theta\,\tau/x)\right\} f(x|\mu,\ \sigma)\,dx, \tag{10}$$

where $w(r)$ is given by (7). Note that for $x = w(r)$ the integrand will vanish. We may write $P(r) = P_1(r) - P_2(r)$ where

$$P_1(r) = \int_{w(r)}^{\infty} x\, f(x|\mu, \sigma)\, dx$$

and

$$P_2(r) = \int_{w(r)}^{\infty} (x/\theta\tau) \ln (1 + r\,\theta\tau/x)\, f(x|\mu,\sigma)\, dx.$$

It is well known (see, for example, Hadley and Whitin [4], p. 446) that since $f(x \mid \mu, \sigma)$ is a normal density,

$$P_1(r) = \sigma\, \phi\left[\frac{w(r) - \mu}{\sigma}\right] + \mu\, \bar{\Phi}\left[\frac{w(r) - \mu}{\sigma}\right], \tag{11}$$

where ϕ is the standard normal density and $\bar{\phi}$ is the complementary cumulative standard normal distribution function. Note that we also used the identity

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma}\, \phi\left[\frac{x - \mu}{\sigma}\right]. \tag{12}$$

Unfortunately, no such direct result is available for determining $P_2(r)$, so that exact computation of $P_2(r)$ requires numerical methods. However, if we use a Taylor series expansion for the $\ln (1 + r\,\theta\,\tau/x)$, the results of the previous section can be applied. The Taylor series expansion for $\ln (1 + t)$ around any point t_o can be shown to be

$$\ln (1 + t) = 1n\ (1 + t_0) + \sum_{n=1}^{\infty} (-1)^{n-1}\, \frac{1}{n} \left[\frac{t - t_0}{1 + t_0}\right]^n \tag{13}$$

where $t = r\,\theta\,\tau/x$.

We have experimented with various values of t_0 and found that the best results were obtained at $t_0 = r\,\theta\,\tau/\mu$. In this case (13) becomes

$$\ln(1 + r\,\theta\,\tau/x) = \ln\left[1 + \frac{r\theta\tau}{\mu}\right] + \sum_{n=1}^{\infty} (-1)^{n-1}\, \frac{1}{n} \left[\frac{\frac{r\theta\tau}{x} - \frac{r\theta\tau}{\mu}}{1 + \frac{r\theta\tau}{\mu}}\right]^n$$

$$= \ln\left[1 + \frac{r\theta\tau}{\mu}\right] + \sum_{n=1}^{\infty} (-1)^{n-1}\, \frac{1}{n} \left[\frac{r\theta\tau}{1 + \frac{r\theta\tau}{\mu}}\right]^n \left[\frac{1}{x} - \frac{1}{\mu}\right]^n$$

$$= \ln\left[1 + \frac{r\theta\tau}{\mu}\right] + \sum_{n=1}^{\infty} (-1)^{n-1}\, \frac{1}{n} \left[\frac{r\theta\tau}{1 + \frac{r\theta\tau}{\mu}}\right]^n$$

$$\sum_{i=0}^{n} \left[\frac{n}{i}\right] \left[\frac{1}{x}\right]^i \left[\frac{-i}{\mu}\right]^{n-i} \tag{14}$$

It is of interest to examine the conditions under which this series will converge. Suppose that $\mu - 3\sigma \leqslant x \leqslant \mu + 3\sigma$. Then

$$\left|\left(\frac{r\theta\tau}{1+\frac{r\theta\tau}{\mu}}\right)^n\left(\frac{1}{x}-\frac{1}{\mu}\right)^n\right| < \left|\frac{\mu}{x}-1\right|^n \leqslant 1 \text{ for } \sigma's \leqslant \frac{\mu}{6}$$

whenever $|x - \mu| \leqslant 3\sigma$, which guarantees convergence of the series for this range of values of x. Note that this agrees with our former requirement in Section 1 that σ should be small in comparison with μ.

Using (14) and truncating after N terms, we now obtain the following approximate expression for $P_2(r)$:

$$\begin{aligned}(15)\quad P_2(r) \approx{}& \frac{\ln\left(1+\frac{r\theta\tau}{\mu}\right)}{\theta\tau}\int_{w(r)}^{\infty} x\, f(x \mid \mu, \sigma)\, dx \\ &+ \frac{1}{\theta\tau}\sum_{n=1}^{N}(-1)^{n-1}\frac{1}{n}\left(\frac{r\theta\tau}{1+\frac{r\theta\tau}{\mu}}\right)^n \int_{w(r)}^{\infty} x \\ &\qquad \sum_{i=0}^{n}\binom{n}{i}\left(\frac{1}{x}\right)^i\left(\frac{-1}{\mu}\right)^{n-i} f(x|\mu, \sigma)\, dx \\ &- \frac{1}{\theta\tau}\left\{\ln\left(1+\frac{r\theta\tau}{\mu}\right) + \sum_{n=1}^{N}(-1)^{n-1}\cdot\frac{1}{n}\cdot\left(\frac{r\theta\tau}{1+\frac{r\theta\tau}{\mu}}\right)^n\left(\frac{-1}{\mu}\right)^n\right\}\cdot P_1(r) \\ &+ \frac{1}{\theta\tau}\sum_{n=1}^{N}(-1)^{n-1}\left(\frac{r\theta\tau}{1+\frac{r\theta\tau}{\mu}}\right)^n\left(\frac{-1}{\mu}\right)^{n-1}\bar{\Phi}\left(\frac{w(r)-\mu}{\sigma}\right) \\ &+ \frac{1}{\theta\tau}\sum_{n=2}^{N}(-1)^{n-1}\frac{1}{n}\left(\frac{r\theta\tau}{1+\frac{r\theta\tau}{\mu}}\right)^n \sum_{i=2}^{n}\binom{n}{i}\left(\frac{-1}{\mu}\right)^{n-1} \\ &\qquad \int_{w(r)}^{\infty}\frac{f(x|\mu, \sigma)}{x^{i-1}}\, dx.\end{aligned}$$

The final equality results from separating out the terms corresponding to $i=0$ and $i=1$ and using the definition of $P_1(r)$, an explicit expression of which is given in (11). The final term can now be approximated by using (3), with the result that

$$\begin{aligned}(16)\quad \int_{w(r)}^{\infty}\frac{f(x|\mu, \sigma)}{x^{i-1}}\, dx \cong{}& \frac{\mu^{i-1}}{\prod_{k=1}^{i-1}(\mu^2 - k\sigma^2)} \\ &\left[\bar{\Phi}\left\{\frac{\mu w(r) - \mu^2 + (i-1)\sigma^2}{\sigma\sqrt{\mu^2-(i-1)\sigma^2}}\right\}\right].\end{aligned}$$

Combining (11), (15), and (16) gives the following computing formula for $P(r)$:

$$P(r) = \left[1 - \frac{1}{\theta\tau} \ln\left[1 + \frac{r\theta\tau}{\mu}\right] + \frac{1}{\theta\tau} \sum_{n=1}^{N} \frac{1}{n} \left[\frac{r\theta\tau}{\mu + r\theta\tau}\right]^n\right] \cdot \left[\sigma\, \phi\left[\frac{w(r) - \mu}{\sigma}\right] + \mu\, \bar{\Phi}\left[\frac{w(r) - \mu}{\sigma}\right]\right] - \frac{\mu}{\theta\tau} \bar{\Phi}\left[\frac{w(r) - \mu}{\sigma}\right] \sum_{n=1}^{N} \left[\frac{r\theta\tau}{\mu + r\theta\tau}\right]^n + \frac{1}{\theta\tau} \sum_{n=2}^{N} \frac{1}{n} \left[\frac{r\theta\tau}{\mu + r\theta\tau}\right]^n \sum_{i=2}^{n} \binom{n}{i} (-\mu)^i \prod_{k=1}^{i-1} \frac{1}{\hat{\mu}_k} \bar{\Phi}\left[\frac{w(r) - \hat{\mu}_{i-1}}{\hat{\sigma}_{i-1}}\right], \tag{17}$$

where

$$\hat{\mu}_k = \frac{\mu^2 - k\sigma^2}{\mu}, \quad \hat{\sigma}_k = \frac{\sigma}{\mu} \sqrt{\mu^2 - k\sigma^2},$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \bar{\Phi}(x) = \int_x^{\infty} \phi(t)\, dt.$$

Although (17) may appear to be fairly complicated, it is completely determined from the knowledge of only the mean and variance of leadtime demand and the standard normal density and distribution functions. Computations were performed to compare the effectiveness of the approximation for $P(r)$ for various combinations of μ, σ, τ, and θ. The exact calculation of $P(r)$ given in (10) was accomplished by numerical integration and the approximation was obtained from (17). Table 2 summarizes the results for $\mu = 10$, $\sigma = 2$, $\tau = 5$, and $N = 6$ for $\theta = 0.1$, 0.3, 0.7, and 1.0. Note that the agreement is quite close throughout and is exact to two decimals for all $10 \leqslant r \leqslant 130$ when $\theta = 1$. Also, the values of $w(r)$ can be seen to change substantially with θ. We also tested the approximation for larger values of μ and found that it gave excellent results as long as μ/σ was relatively large (say 6 or more). It is worth pointing out that although (17) is fairly complicated, it can be determined far more quickly for moderate values of N than can (10) by numerical integration. This can be extremely important when the expected number of shortages incurred during leadtime must be computed for a range of values of r.

REFERENCES

[1] DeGroot, M. H., *Probability and Statistics* (Addison Wesley, Reading, Mass, 1975.)

[2] Emmons, H., "The Optimal Use of Radioactive Pharmaceuticals," Unpublished Ph.D. Thesis. The Johns Hopkins University (1968).

[3] Ghare, P. M., and G. F. Schrader, "A Model For Exponentially Decaying Inventory," *Journal of Industrial Engineering 14*, 238-243, (1963).

[4] Hadley, G., and T. M. Whitin, *Analysis of Inventory Systems* (Prentice Hall, Englewood Cliffs, New Jersey, 1963).

TABLE 2 — *A Comparison Between Exact and Approximate Expression for Expected Shortage for* $\mu = 10$, $\sigma = 2$, $N = 6$, $\tau = 5$.

r	$\theta = 0.1$			$\theta = 0.3$			$\theta = 0.7$			$\theta = 1.0$		
	$W(r)$	Exact	App	$W(r)$	Exact	App	$W(r)$	Exact	App	$W(r)$	Exact	App
10	7.71	2.03	2.05	4.31	3.94	3.94	1.09	5.74	5.74	0.34	6.44	6.44
15	11.56	0.20	0.22	6.46	2.22	2.27	1.64	4.80	4.80	0.51	5.75	5.75
20	15.42	0.00	-.00	8.62	0.97	1.02	2.18	4.10	4.10	0.68	5.24	5.24
25	19.26	0.00	0.00	10.78	0.26	0.35	2.73	3.54	3.54	0.85	4.83	4.83
30		0.00	0.00	12.93	0.03	0.01	3.27	3.07	3.07	1.02	4.50	4.50
35		0.00	0.00	15.08	0.00	0.00	3.81	2.67	2.67	1.19	4.20	4.20
40		0.00	0.00		0.00	−0.00	4.36	2.31	2.32	1.36	3.95	3.95
45		0.00	0.00		0.00	0.00	4.90	2.00	2.00	1.53	3.72	3.72
50		0.00	0.00		0.00	0.00	5.45	1.72	1.73	1.70	3.52	3.52
55		0.00	0.00		0.00	0.00	5.99	1.46	1.49	1.87	3.34	3.34
60		0.00	0.00		0.00	0.00	6.54	1.23	1.28	2.04	3.17	3.17
65		0.00	0.00		0.00	0.00	7.08	1.02	1.09	2.21	3.01	3.01
70		0.00	0.00		0.00	0.00	7.63	0.83	0.89	2.37	2.87	2.87
75		0.00	0.00		0.00	0.00	8.17	0.66	0.78	2.54	2.74	2.74
80		0.00	0.00		0.00	0.00	8.72	0.51	0.50	2.71	2.61	2.61
85		0.00	0.00		0.00	0.00	9.26	0.38	0.52	2.89	2.49	2.49
90		0.00	0.00		0.00	0.00	9.81	0.28	0.33	3.05	2.38	2.38
95		0.00	0.00		0.00	0.00	10.35	0.19	0.32	3.22	2.28	2.28
100		0.00	0.00		0.00	0.00	10.90	0.13	0.24	3.39	2.18	2.18
105		0.00	0.00		0.00	0.00				3.56	2.08	2.08
110		0.00	0.00		0.00	0.00				3.73	1.99	1.99
115		0.00	0.00		0.00	0.00				3.90	1.90	1.90
120		0.00	0.00		0.00	0.00				4.07	1.82	1.82
125		0.00	0.00		0.00	0.00				4.24	1.74	1.74
130		0.00	0.00		0.00	0.00				4.40	1.66	1.66

[5] Nahmias, S., "Optimal Ordering Policies For Perishable Inventory-II," Operations Research, *23*, 735-749, (1975).

[6] Nahmias, S., G. Levine, and Y. Choy, "A Cost Benefit Inventory Management System For Radioactive Pharmaceuticals," Proceedings of the Sixth Annual Modelling and Simulation Conference, 1247-1251, (1975).

ALL-INTEGER LINEAR PROGRAMMING — A NEW APPROACH VIA DYNAMIC PROGRAMMING

Leon Cooper and Mary W. Cooper

Department of Operations Research & Engineering Mangagement
School of Engineering & Applied Science
Southern Methodist University
Dallas, Texas

ABSTRACT

An exact method for solving all-integer linear-programming problems is presented. Dynamic-programming methodology is used to search efficiently candidate hyperplanes for the optimal feasible integer solution. The explosive storage requirements for high-dimensional dynamic programming are avoided by the development of an analytic representation of the optimal allocation at each stage. Computational results for problems of small to moderate size are also presented.

INTRODUCTION

The problem to be considered in this paper is the all-integer linear-programming problem, which is as follows:

$$\begin{aligned} &\max z = c'x, \\ &Ax \leqslant b, \text{ and} \\ &x \geqslant 0, \text{ integer.} \end{aligned} \tag{1}$$

In (1), c and x are n-component vectors and c has integral nonnegative components. A is an $m \times n$ matrix and b is an m-component vector. Furthermore, we assume that z is bounded and that there exists at least one feasible integer point in the convex set $S = \{x | Ax \leqslant b,\ x \geqslant 0\}$.

In an earlier paper by one of the authors [1], a precursor of the formulation of the method proposed in this paper was made. However, it was not efficient from the point of view of computer storage. A basic reformulation of the method is made in this paper which changes the previous algorithm very significantly. In the present paper, the basic theory to make the algorithm computationally efficient is presented. The new algorithm is presented along with some computational results. It should be emphasized that this algorithm is exact and suffers from no numerical problems of convergence.

The assumption that the components of c are nonnegative is equivalent to the statement that each of the separable terms of the objective function is nondecreasing. This is required by the dynamic programming approach we shall use. This is not a limitation in generality, since if any of the c_j, $j = 1, 2, \ldots, n$, are negative, a simple transformation may be made to convert the negative coefficients. This will be discussed in a subsequent section.

It will be noted that if any attempt were made to solve the problem given by (1) by dynamic programming it would rapidly run into the "curse of dimensionality," for $m \geqslant 3$ or 4 and the storage requirements on a computer would be not only prohibitive but nonexistent. The method proposed here avoids this explosive increase of memory requirement when the dimensionality is high. We will, however, exploit the use of dynamic-programming methodology, as will be seen subsequently.

2. GENERAL DESCRIPTION OF THE ALGORITHM

The general idea of the proposed algorithm is to search candidate hyperplanes for lattice points. This proceeds as follows. If we remove the integer requirement from (1), the relaxed problem can be solved as a linear-programming problem. Suppose the optimal value of the objective function for this relaxed problem is z°. In addition, let the optimal value of the objective function for (1) be designated z^*. It is clear that $z^* \leqslant z^\circ$. The basic idea behind the hyperplane search algorithm we propose is to start at the linear-programming solution and search the hyperplane $c'x = [z^\circ]$ (where $[\alpha]$ indicates the greatest integer less than or equal to α) to see whether or not it contains any feasible lattice points. If it does, we are done. If it does not, we move the hyperplane in a direction parallel to itself and then search the hyperplane $c'x = [z^\circ] - 1$. Since c was assumed to have integral components, $c'x$ must be an integer if x is to have all integral components. If the hyperplane $c'x = [z^\circ] - 1$ contains at least one feasible lattice point, we are done. If it does not, we continue the process. This procedure is clearly finite. Since it was assumed that S was nonempty and contained at least one lattice point, we must eventually find it. Let us now describe this algorithm or class of algorithms more precisely.

We summarize the notation we will use:

z^* = optimal value of objective function in (1),

z° = optimal solution to linear-programming problem derived from (1),

$z_k = z^\circ - k,\ k = 0, 1, 2, \ldots,$

$S = \{x \mid Ax \leqslant b,\ x \geqslant 0\},$

ℓ_j = lower bounds on x_j, i.e., $x_j \geqslant \ell_j,\ j = 1, 2, \ldots, n$, and

u_j = upper bounds on x_j, i.e., $x_j \leqslant u_j,\ j = 1, 2, \ldots, n.$

HYPERPLANE SEARCH ALGORITHM

1. Solve the linear programming problem

$$\max z = c'x$$
$$Ax \leqslant b, \text{and}$$
$$x \geqslant 0. \tag{2}$$

If x°, the optimal solution to (2), satisfies the requirement of being all integer, we are done. Otherwise, proceed to step 2.

2. Determine lower bounds ℓ_j and upper bounds u_j for each variable x_j. We then have

$$\ell_j \leqslant x_j \leqslant u_j,$$
$$x_j \text{ integer}, \quad j = 1, 2, \ldots, n. \tag{3}$$

3. Find all combinations of x_j, $j = 1, 2, \ldots, n$, which satisfy

$$c'x = z_k$$

$$\ell_j \leqslant x_j \leqslant u_j, \quad j = 1, 2, \ldots, n. \tag{4}$$

4. If no integer valued vector x^k can be found, increase k by 1, i.e., decrease z_k by 1 and return to step 3 (or in some cases step 2 — see discussion below). If at least one x^k is all integer, go to step 5.

5. If at least one $x^k \in S$, we are done. If for all x^k, $x^k \in S$, decrease z_k by 1 and return to step 3 (or in some cases step 2 — see discussion below).

Let us now consider each step of the above algorithm in greater detail. First, we may note that since the set S is nonempty, is bounded, and contains at least one integer point, the finiteness of the algorithm is guaranteed. How efficient such an algorithm can be depends very strongly on how step 3 is carried out. Let us consider each step in turn.

Step 1 requires little comment. Any simplex code can be used to solve the linear-programming problem given by (2).

Step 2 should be carried out in the most convenient fashion. Lower bounds of zero on the x_j can always be used. If upper bounds can be easily determined from the physical intepretation of the variables, they should be used. Linear-programming could also be used, if necessary, to determine both the lower and upper bounds on the x_j by solving the problems

$$\begin{array}{ll} \min x_j, & \max x_j, \\ Ax \leqslant b, & Ax \leqslant b, \\ c'x = z_k, & c'x = z_k, \\ x \geqslant 0. & x \geqslant 0. \end{array} \tag{5}$$

These problems could be solved once (initially) to get bounds, or periodically, as steps 4 and 5 of the algorithm indicate. In any case, lower and upper bounds can be found.

Steps 4 are 5 and self-evident and require no extensive comment except to note that this is the only place the structural constraints of (1) enter the problem, with the possible exception of the determination of bounds. As we shall see, the determination of the solution x^k in step 3 does not explicitly depend upon the constraints $Ax \leqslant b$.

3. HYPERPLANE SEARCH BY DYNAMIC PROGRAMMING

We now consider how we may use a dynamic programming formulation and method of solution to deal with step 3 of the hyperplane search algorithm. This amounts to finding all combinations of x_j which satisfy (4). We may formulate this problem as

$$\max z = c'x \tag{6}$$

$$c'x = z_k,$$

$$c_j > 0, \quad j = 1, 2, \ldots, n,$$

$$0 \leqslant \ell_j \leqslant x_j \leqslant u_j, \quad j = 1, 2, \ldots, n, \text{ and}$$

$$x_j \text{ integer}, \quad j = 1, 2, \ldots, n.$$

The fact that we already know the maximum value of z for this subproblem, viz, z_k, in no way invalidates (6) as a meaningful problem, since what we are seeking is whether or not there exists a set of values x satisfying the constraints of (6). It will be noted, as mentioned previously, that we need to assume that all $c_j > 0$. Let us demonstrate that this does not result in any loss of generality.

LEMMA 1: An integer programming problem of the form given by (6), except that not all of the objective function coefficients are positive, may be transformed into an equivalent problem all of whose coefficients are positive.

PROOF: The proof is by construction. Suppose we have the problem

$$\begin{aligned} &\max z = c'x, \\ &c'x = z_k, \\ &\ell \leqslant x \leqslant u, \end{aligned} \tag{7}$$

where c is an integer vector, but not all $c_j > 0$. If $c_j > 0$, let $\hat{x}_j = x_j$ and $\hat{c}_j = c_j$, and if $c_j < 0$, left $\hat{x}_j = u_j - x_j$ and $\hat{c}_j = -c_j$. Further, let $P = \{j | c_j > 0\}$ and $N = \{j | c_j < 0\}$. We may then rewrite the objective function of (7) as

$$\begin{aligned} \max z &= \sum_{j \in P} c_j x_j + \sum_{j \in N} c_j x_j \\ &= \sum_{j \in P} c_j x_j + \sum_{j \in N} c_j (u_j - x_j) \\ &= \hat{c}'x + \sum_{j \in N} \hat{c}_j u_j. \end{aligned} \tag{8}$$

Hence, instead of solving (7) directly, we may solve the equivalent problem

$$\begin{aligned} &\max z = c'x, \\ &\hat{c}'\hat{x} = z_k + \sum_{j \in N} \hat{c}_j u_j = \hat{z}_k, \ \hat{\ell} \leqslant \hat{x} \leqslant \hat{u}, \end{aligned} \tag{9}$$

where

$$\begin{aligned} &\hat{\ell}_j = \ell_j \quad \text{if } c_j > 0, \\ &\hat{u}_j = u_j \end{aligned} \tag{10}$$

and

$$\begin{aligned} &\ell_j = 0 \qquad \text{if } c_j < 0, \\ &\hat{u}_j = u_j - \ell_j \end{aligned} \tag{11}$$

and now the equivalent problem (9) has all $c_j > 0$.

The problem given in (6) can be solved by the use of dynamic programming. Since the objective function and the single structural constraint are separable and nondecreasing ($c_j > 0$) functions, the sufficient conditions for a solution by dynamic-programming are satisfied (see Ref. [2]). By applying the principle of optimality the dynamic-programming solution to (6) is easily obtained. The optimal return functions $g_s(\cdot)$ are given by the following recursion relations:

$$g_1(\lambda) = \max_{x_1 = \delta_1} c_1 x_1 = \begin{cases} \lambda, & \lambda = c_1 t, \ t = 0, 1, \ldots, u_1, \\ -\infty & \text{otherwise.} \end{cases} \tag{12}$$

where

$$\delta_1 = \begin{cases} \dfrac{\lambda}{c_1}, & \lambda = c_1 t,\ t = 0, 1, \ldots, u_1, \\ \text{undefined, otherwise;} \end{cases} \tag{13}$$

$$g_s(\lambda) = \max_{l_s \leq x_s \leq \delta_s(\lambda)} [c_s x_x + g_{s-1}(\lambda - c_s x_s)], \quad \begin{matrix} s = 2, 3, \ldots, n, \\ \lambda = 0, 1, \ldots, \Lambda_s, \end{matrix} \tag{14}$$

$$\delta_s(\lambda) = \min\left[u_s, \left[\frac{\lambda}{c_s}\right]\right] \tag{15}$$

$$\Lambda_s = \sum_{j=1}^{s} c_j u_j. \tag{16}$$

The usual dynamic-programming approach would be to calculate

$$g_s(\lambda) \text{ and } x_s^*(\lambda), \qquad \lambda = 0, 1, \ldots, \Lambda_s,$$

for $s = 1, 2, \ldots, n-1$, where $x_s^*(\lambda)$ is the value of x_s which produced $g_s(\lambda)$ for each value of λ. Finally, we would calculate $g_n(z_k)$ and $x_n^*(z_k)$, assuming a solution exists. We would then subtract $c_n x_n^*$ from z_k and then find, corresponding to $\lambda = z_k - c_n x_n^*$, in the tabulation of $x_{n-1}^*(\lambda)$, the value of $x_{n-1}^*(z_k - c_n x_n^*)$ which gave rise to $g_{n-1}(z_k - c_n x_n^*)$. This backwards process would yield, successively, $x_n^*, x_{n-1}^*, \ldots, x_1^*$.

The principle difficulty with this approach is the storage requirement, which, while it is orders of magnitude less than that for the simple-minded approach of using a state variable for each constraint of (1), still is quite considerable. For each variable a vector $x_s^*(\lambda)$ must be stored. Furthermore, there are often many alternate optimal values of $x_s^*(\lambda)$ which maximize $c_s x_s + g_{n-1}(\lambda - c_s x_s)$ and they must all be stored. Hence $x_s^*(\lambda)$ is actually a matrix, say of average dimension $\approx \left(\dfrac{u_s}{2} \times \Lambda_s\right)$. Hence the total amount of storage required is approximately $\sum_{s=1}^{n-1} \dfrac{u_s \Lambda_s}{2}$. For example, if all c_j were 5, all $u_j = 10$, and n = 100, we would require about 1,260,000 words of computer storage. There are ways to minimize this, but nevertheless, with increasing n, the storage problem becomes significant.

In the following section, a set of equations will be derived by the use of simple combinatorial theory, which will give explicit formulae for $x_s^*(\lambda)$ for any λ, and hence the need for a complete tabulation of $x_s^*(\lambda)$ will be eliminated. Indeed $g_n(\lambda)$ need never be explicitly calculated. The reduction in storage is drastic and renders the hyperplane method just described of practical use. The entire calculation process will be reduced to calculating $x_n^*(z_k)$, $x_{n-1}^*(z_k - c_n x_n^*), \ldots, x_1^*(z_k - \sum_{s=2}^{n} c_s x_s^*)$ directly.

4. DERIVATION OF EQUATIONS FOR OPTIMAL SOLUTION

The derivations that follow are concerned with deriving a set of expressions that yield $x_s^*(\lambda)$ by applying the recursion relations (12) to (16) to the problem

$$\max z = \sum_{j=1}^{n} c_j x_j, \qquad (c_1 = 1)$$

$$\sum_{j=1}^{n} c_j x_j = z_k,$$

$$\ell_j \leqslant x_j \leqslant u_j, \quad j = 1, 2, \ldots, n,$$

(17) $$c_j \text{ integer}, \quad j = 1, 2, \ldots, n.$$

It will be noted that we have assumed that $c_1 = 1$. This is not absolutely necessary but it does result in a drastic simplification of the expressions for $x_s^*(\lambda)$. There is no loss in generality in doing so, as the following lemma shows.

LEMMA 2: Given the problem

$$\max z = c'x,$$

$$c'x = z_k,$$

$$\ell \leqslant x \leqslant u,$$

(18) $$c > \bar{1}, \text{ integer},$$

a problem satisfying the conditions of (17) can be derived which contains the solution to (18) as a subset.

PROOF: The proof is by construction. If we add a variable $x_o = 0$ to the vector x then the following problem has a solution which contains the solutions to (18) as a subset.

$$\max z = x_o + \sum_{j=1}^{n} c_j x_j,$$

$$x_o + \sum_{j=1}^{n} c_j x_j = z_k,$$

$$\ell \leqslant x \leqslant u,$$

$$x_o = 0,$$

(19) $$c \text{ integer}.$$

By adding the constraint $x_o = 0$ to $Ax \leqslant b$ of the original integer linear programming, we force x_o to be zero and hence have a problem which has as a subset the solution to (18).

In what follows then, we always assume that $c_1 = 1$. We shall present the equations for optimal allocations $x_s^*(\lambda)$, $s = 1, 2, \ldots, n$ and proofs of their validity in the next set of lemmas. Without any loss of generality, we shall assume that $\ell_s = 0$, $s = 1, 2, \ldots, n$. That this is so is obvious, since if $\ell \leqslant x \leqslant \hat{u}$ and if $y = x - \ell$, $o \leqslant y \leqslant \hat{u} - \ell = u$. Therefore, the problem for which we are developing a set of expressions for $x_s^*(\lambda)$ is:

$$\max z = \sum_{j=1}^{n} c_j x_j \qquad (c_1 = 1),$$

$$\sum_{j=1}^{n} c_j x_j = z_k,$$

$$0 \leqslant x_j \leqslant u_j, \qquad j = 1, 2, \ldots, n,$$

(20) $$c_j \text{ integer}, \quad j = 1, 2, \ldots, n.$$

LEMMA 3: $x_1^*(\lambda) = \lambda, \ 0 \leqslant \lambda \leqslant u_1$.

PROOF: The dynamic-programming solution, which results from the application of the principle of optimality to the first stage, is:

$$g_1(\lambda) = \max_{x_1 = \lambda} x_1 = \lambda, \qquad 0 \leqslant \lambda \leqslant u_1, \ \text{integer}. \tag{21}$$

This must be the case because, in the backward allocation process, the first stage is reached last. If $c_1 = 1$ and λ is left to allocate, then clearly $x_1^*(\lambda) = \lambda$.

LEMMA 4: $x_s^*(\lambda) = 0$, $\lambda \leqslant c_s - 1 \leqslant \Lambda_{s-1}$, $s = 2, 3, \ldots, n$.

PROOF: The dynamic programming recursion relations for $2 \leqslant s \leqslant n$ are as follows

$$g_s(\lambda) = \max_{0 \leqslant x_s \leqslant \delta_s(\lambda)} [c_s x_s + g_{s-1}(\lambda - c_s x_s)]. \tag{22}$$

Let $\lambda - c_s x_s = \xi \geqslant 0$. If $\lambda - c_s x_s < 0$, $g_{s-1}(\xi)$ does not exist.

CASE 1 $\lambda = c_s - 1$:
We have

$$c_s - 1 - c_s x_s = \xi, \ c_s \geqslant 1,$$

and

$$c_s(1 - x_s) - 1 = \xi, \ \ c_s \geqslant 1.$$

If

$$x_s = 0, \ \xi = c_s - 1, \ g_{s-1}(\xi) = \xi = c_s - 1 \leqslant \Lambda_{s-1},$$

and

$$x_s^*(\lambda) = 0.$$

If $x_s \geqslant 1$, $\xi > 0$. Hence g_{s-1} does not exist. $x_s^*(\lambda) = 0$ is the only solution.

CASE 2, $\lambda < c_s - 1$:
Let

$$\lambda = c_s - t, \ \ 1 < t \leqslant c_s,$$

then

$$\lambda - c_s x_s = c_s - t - c_s x_s = \xi \geqslant 0,$$

and

$$c_s(1 - x_s) - t = \xi.$$

If

$$x_s = 0, \ \xi = c_s - t \geqslant 0, \ g_{s-1}(\xi) = c_s - t \leqslant \Lambda_{s-1},$$

and

$$x_s^*(\lambda) = 0.$$

If $x_s \geqslant 1$, $\xi < 0$. Hence g_{s-1} does not exist. $x_s^*(\lambda) = 0$ is the only solution.
We have therefore proved that:

$$x_s^*(\lambda) = 0, \ \ \lambda \leqslant c_s - 1 \leqslant \Lambda_{s-1}, \ \ s = 2, 3, \ldots, n$$

LEMMA 5: $x_s^*(\lambda) = 0, 1, 2, \ldots, \delta_s(\lambda)$, $c_s \leqslant \lambda \leqslant \Lambda_{s-1}$, where

$$\delta_s(\lambda) = \min\left(\left[\frac{\lambda}{c_s}\right], u_s\right) \text{ for } s = 2, 3, \ldots, n.$$

PROOF: The dynamic-programming recursion relations for $2 \leqslant s \leqslant n$ are:

$$g_s(\lambda) = \max_{0 \leqslant x_s \leqslant \delta_s(\lambda)} [c_s x_s + g_{s-1}(\lambda - c_s x_s)]. \tag{23}$$

Since $\lambda \geqslant c_s$ and $\delta_s(\lambda) = \min\left(\left[\frac{\lambda}{c_s}\right], u_s\right)$, then $\lambda - c_s x_s \geqslant 0$. Furthermore, $\lambda \leqslant \Lambda_{s-1}$ and $g_{s-1}(\cdot)$ and $x_{s-1}(\cdot)$ have been defined for all values $0 \leqslant \xi \leqslant \Lambda_{s-1}$. Hence, it is clear that $g_s(\lambda) = \lambda$ and hence a value may be assigned for each x_s, $0 \leqslant x_s \leqslant \delta_s(\lambda)$, corresponding to each term in brackets in (23). Therefore,

$$x_s^*(\lambda) = 0, 1, 2, \ldots, \delta_s(\lambda), \qquad \begin{matrix} c_s \leqslant \lambda \leqslant \Lambda_{s-1}, \\ s = 2, 3, \ldots, n. \end{matrix}$$

LEMMA 6:

$$x_s^*(\lambda) = \left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right] + 1, \left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right] + 2, \ldots, v_{s-1}, v_s; \; \Lambda_{s-1} < \lambda \leqslant \Lambda_s$$

where

$$v_s = u_s - t^*, \quad \lambda \geqslant [c_s(u_s - t^*)],$$

and

$$t^* = \max_t [c_s(u_s - t) - \lambda] \leqslant 0, \; t = 0, 1, 2, \ldots, \; s = 2, 3, \ldots, n.$$

PROOF: The dynamic programming recursion relations are

$$g_s(\lambda) = \max_{0 \leqslant x_s \leqslant \delta_s(\lambda)} [c_s x_s + g_{s-1}(\lambda - c_s x_s)], \quad \Lambda_{s-1} < \lambda \leqslant \Lambda_s. \tag{24}$$

We may rewrite (24) as

$$g_s(\lambda) = \max[0 + g_{s-1}(\lambda), \; c_s + g_{s-1}(\lambda - c_s), \; \begin{matrix} 2c_s + g_{s-1}(\lambda - 2c_s), \ldots, \\ \delta_s(\lambda)\, c_s + g_{s-1}(\lambda - \delta_s(\lambda)\, c_s)] \end{matrix}.$$

Since $\lambda > \Lambda_{s-1}$ and $g_{s-1}(\lambda)$ for $\lambda > \Lambda_{s-1}$ is undefined, it is clear that $x_s^*(\lambda) \neq 0$. Therefore, the minimum value of $x_s^*(\lambda) > 0$. To determine the minimum value we note that

$$\lambda - c_s x_s \leqslant \Lambda_{s-1} + 1;$$

Therefore,

$$x_s \geqslant \frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}. \tag{25}$$

However, x_s must be an integer. Hence we know that

$$x_s \geqslant \left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right], \tag{26}$$

and

$$x_s \neq 0. \tag{27}$$

The above facts can be combined by means of the following. Since $x_s^* \neq 0$, $\lambda \geqslant \Lambda_{s-1} + 1 + c_s$. If we substitute $\lambda = \Lambda_{s-1} + 1 + c_s$ into (26) we have

$$x_s \geqslant \frac{\Lambda_{s-1} + 1 + c_s - (\Lambda_{s-1} + 1)}{c_s} = 1.$$

However, a value of 1 under this condition is not possible since

$$g_{s-1}(\Lambda_{s-1} + 1 + c_s - c_s) = g_{s-1}(\Lambda_{s-1} + 1),$$

and $g_{s-1}(\Lambda_{s-1} + 1)$ is undefined. More generally, if $\lambda = \Lambda_{s-1} + q$, then

$$g_s(\lambda) = \max[0 + g_{s-1}(\Lambda_{s-1} + q),\ c_s g_{s-1}(\Lambda_{s-1} + q - c_s),\ \ldots,$$
$$\delta_s(\lambda)\ c_s + g_{s-1}(\Lambda_{s-1} + q - \delta_s(\lambda)\ c_s)$$

will contain terms for which $g_{s-1}(\lambda)$ is not defined. This will occur precisely for those terms for which $q > c_s$. This leads to the minimum value for x_s^* as

$$\left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right] + 1. \tag{28}$$

Larger values of x_s^* will obviously be permitted, since $\lambda - c_s x_s$ will decrease as x_s increases, and hence values of $g_{s-1}(\cdot)$ will exist for these arguments. However, there is an upper bound on x_s^*. This will be called v_s and is derived as follows. If $\lambda \geqslant c_s u_s$, then clearly the largest value of x_s is u_s since

$$g_s(c_s u_s) = \max[0 + g_{s-1}(c_s u_s),\ c_s + g_{s-1}(c_s u_s - c_s),\ \ldots,\ c_s u_s + g_{s-1}(0)]$$

It is clear that no entry beyond the last is possible and this corresponds to $x_s = u_s$. However, if $\lambda < c_s u_s$, we wish to find the largest value of x_s compatible with that value of λ. We recall that when $\lambda \geqslant c_s u_s$, the largest value of $x_s = u_s$. Let us now suppose that

$$\lambda \geqslant c_s v_s = c_s(u_s - t^*) \tag{29}$$

In order to make v_s as large as possible in (29) it is clear that

$$t^* = \max_t[c_s(u_s - t) - \lambda] \leqslant 0,\quad t = 0, 1, 2, \ldots,$$

i.e., t^* is the minimum value of t such that $\lambda \geqslant c_s v_s = c_s(u_s - t^*)$. Then $v_s = u_s - t^*$. It is seen that when $\lambda \geqslant c_s u_s$, $t^* = 0$.

We have not considered the possibility that v_s may be less than $\left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right] + 1$. The significance of this is treated in Lemmas 7 and 8.

Lemmas 3, 4, 5, 6 together constitute the following theorem which gives the formulae to be used in the backwards recursion for a solution by dynamic programming of the problem stated in (20). The optimal values of $x_s^*(\lambda)$ for any λ and for all s can be calculated from these equations.

THEOREM 1: The optimal returns $x_s^*(\lambda)$ for any λ and all s which constitute the solution to (20) are included in the following:

$$x_1^*(\lambda) = \lambda, \quad 0 \leqslant \lambda \leqslant u_1, \quad \lambda \text{ integer}, \tag{30}$$

$$0, \ \lambda \leqslant c_s - 1 \leqslant \Lambda_{s-1}, \qquad s = 2, 3, \ldots, n, \tag{31}$$

$$0, 1, 2, \ldots, \delta_s(\lambda), \qquad c_s \leqslant \lambda \leqslant \Lambda_{s-1}, \qquad s = 2, 3, \ldots, n, \tag{32}$$

$$\left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right] + 1, \left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right] + 2, \ldots, v_{s-1}, v_s, \tag{33}$$

$$x_s^*(\lambda) = \begin{cases} \\ \\ \text{where } v_s = u_s - t^*, \ \lambda \geqslant [c_s(u_s - t^*)], \\ t^* = \max_t [c_s(u_s - t) - \lambda] \leqslant 0, \quad t = 0, 1, 2, \ldots, \quad s = 2, 3, \ldots, n, \\ \text{and} \\ \text{undefined if } \lambda > \Lambda_s, \qquad s = 2, 3, \ldots, n. \end{cases} \tag{34}$$

For the proof of Theorem 1, see Lemmas 3 to 6.

It should be noted that the statement of Theorem 1 implies that the relations given by (30) to (34) also include values which are *not* optimal returns. This is so because of necessity; the dynamic-programming solution generates negative infinite returns for some values of λ (see Ref. [1]), i.e., some values of $g_s(\lambda)$ are not defined. The next two lemmas deal with this situation.

LEMMA 7: If

$$v_s < \left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right] + 1, \ \Lambda_{s-1} < \lambda \leqslant \Lambda_s, \text{ then } c_s > \Lambda_{s-1} + 1.$$

PROOF: By hypothesis, we have that

$$v_s < \left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right] + 1. \tag{35}$$

Since $[\alpha] \leqslant \alpha$, we can remove the integer requirement of

$$\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}$$

in (35) and strengthen the inequality. Hence we have

$$v_s < \frac{\lambda - (\Lambda_{s-1} + 1)}{c_s} + 1. \tag{36}$$

Since $c_s > 0$, we can rewrite (36) as

$$\lambda + c_s - (\Lambda_{s-1} + 1) > c_s v_s \tag{37}$$

By definition, when $\Lambda_{s-1} < \lambda \leqslant \Lambda_s$ we know that

$$v_s = u_s \text{ or } \lambda < c_s v_s + c_s. \tag{38}$$

The inequalities (37) and relations (38) together imply that

$$c_s - (\Lambda_{s-1} + 1) > 0. \tag{39}$$

Rewriting (39) we have

$$c_s > \Lambda_{s-1} + 1,$$

which was to be shown.

LEMMA 8: If $c_s > \Lambda_{s-1} + 1$, then $g_s(\lambda) = -\infty$ for $\lambda = \Lambda_{s-1} + 1 + pc_s$, $p = 0, 1, \ldots,$ and $x_s^*(\lambda)$ is undefined.

PROOF: The dynamic programming recursion relations are

$$g_s(\lambda) = \max_{0 \leqslant x_s \leqslant \delta_s(\lambda)} [c_s x_s + g_{s-1}(\lambda - c_s x_s)], \qquad s = 2, 3, \ldots, n.$$

Consider $g_s(\lambda)$, $\lambda = \Lambda_{s-1} + 1 + pc_s$, $p = 0, 1, \ldots$. We then have

$$g_s(\Lambda_{s-1} + 1 + pc_s) = \max[0 + g_{s-1}(\Lambda_{s-1} + 1 + pc_s),\ c_s + g_{s-1}(\Lambda_{s-1} + 1 + pc_s - c_s),$$
$$\ldots,\ c_s\delta_s(\lambda) + g_{s-1}(\Lambda_{s-1} + 1 + pc_s - c_s\delta_s(\lambda))],$$

(40) $$\text{where } \delta_s(\lambda) = \min\left[u_s,\ \left[\frac{\Lambda_{s-1} + 1 + pc_s}{c_s}\right]\right].$$

Consider the terms in the brackets in (40). The first term exceeds the limit Λ_{s-1} of the function $g_{s-1}(\cdot)$ for all values of p. Since $c_s > \Lambda_{s-1} + 1$ by hypothesis, all arguments of $g_{s-1}(\cdot)$ in subsequent terms will either exceed the limit Λ_{s-1} or be undefined if the argument is negative. This will depend upon the relative magnitudes of p and x_s. However, in no case can the value be both less than or equal to Λ_{s-1} and nonnegative. This follows because the argument of g_{s-1} is

(41) $$\Lambda_{s-1} + 1 + c_s(p - x_s)$$

CASE 1, $p < x_s$:

It then follows that $p - x_s < 0$. Since $c_s > \Lambda_{s-1} + 1$, $\Lambda_{s-1} + 1 + c_s(p - x_s) < 0$, and negative arguments for $g_{s-1}(\cdot)$ are not defined.

CASE 2, $p = x_s$:

It then follows that (41) reduces to $\Lambda_{s-1} + 1$, which exceeds the limit of the arguments for $g_{s-1}(\cdot)$.

Case 3, $p > x_s$:

$$c_s(p - x_s) > 0.$$

Therefore $\Lambda_{s-1} + 1 + c_s(p - x_s)$ exceeds the limit Λ_{s-1} allowable for $g_{s-1}(\cdot)$.

This completes the proof.

The import of Lemmas 7 and 8 is to settle the question raised at the end of Lemma 6 on what was the significance of $v_s < \left[\frac{\lambda - (\Lambda_{s-1} + 1)}{c_s}\right] + 1$ for $\Lambda_{s-1} < \lambda \leqslant \Lambda_s$. By Lemma 7, this fact implies that $c_s > \Lambda_{s-1} + 1$, and Lemma 8 tells us that if this latter fact is true then $x_s^*(\lambda)$ is undefined. Hence, the backwards recursion may be discontinued for the set of values x_s, being tested with the current value of z_k, and the next set of x_s values may be tested.

The significance of Theorem 1 and Lemmas 7 and 8 is that the entire backwards recursion for the optimal values $x_n^*(z_k)$, $x_{n-1}^*(z_k - c_n x_n^*)$, ..., $x_1^*(z_k - \sum_{s=2}^{n} c_s x_s^*)$ may be calculated, given any value of z_k, from the equations (30) to (34) in Theorem 1 without ever carrying out the forward calculation and storage of lengthy tables. Furthermore, Lemmas 7 and 8 tell us that in the course of the calculation if we encounter for some $s = q$ that $c_q > \Lambda_{q-1} + 1$, we may terminate the calculation for the current z_k and proceed to begin again with $z_k - 1$.

In the next section we shall give two examples of the use of the theory we have just developed.

5. HYPERPLANE SEARCH ALGORITHM — TWO EXAMPLES

Example 1: *Max* $z = x_1 + 3x_2 + 4x_3 + 6x_4$,

$$2x_1 + 3x_2 + 6x_3 + 4x_4 \leqslant 23,$$

$$5x_1 + 4x_2 + 2x_3 + x_4 \leqslant 20,$$

$$x_j \geqslant 0, \text{ integer}, \quad j = 1, 2, 3, 4. \tag{41}$$

First we solve (41) as a linear-programming problem, ignoring the integrality requirement. If we do so, we obtain the solution $x_1^* = 0$, $x_2^* = 2.056$, $x_3^* = 1.111$ and $z^* = 9.5$. Hence $z^\circ \leqslant 9$. Let us now apply the hyperplane search algorithm equations of Theorem 1.

It is readily seen from the constraints of (41) that the variables are bounded as follows:

$$0 \leqslant x_1 \leqslant 7, \quad 0 \leqslant x_2 \leqslant 2, \quad 0 \leqslant x_3 \leqslant 2, \text{ and } 0 \leqslant x_4 \leqslant 3.$$

However, if we solve a linear programming problem in which we maximize each variable in turn, we find the following closer bounds:

$$0 \leqslant x_1 \leqslant 5, \quad 0 \leqslant x_2 \leqslant 2, \quad 0 \leqslant x_3 \leqslant 1, \text{ and } 0 \leqslant x_4 \leqslant 0.$$

Therefore, we have

$$\Lambda_1 = 5, \quad \Lambda_2 = 11, \quad \Lambda_3 = 14, \text{ and } \Lambda_4 = 14.$$

We know from the bounds that $x_4^*(9) = 0$. For $\lambda = 9$ we have that $c_3 = 3 \leqslant 9 \leqslant 14 = \Lambda_3$. Therefore, we have that $x_3^*(9) = 0, 1$. Since the LP solution gave $x_3 = 1.111$, we try $x_3^* = 1$. Then $\lambda = 6$, and $c_2 = 3 \leqslant 6 \leqslant 11 = \Lambda_2$. Hence we have that $x_2^*(6) = 0, 1, 2$. Since the LP solution had $x_2 = 2.056$, we try $x_2^* = 2$, which leaves $\lambda = 0$ and $x_1 = 0$. We now have as a candidate solution $x_1 = 0$, $x_2 = 2$, $x_3 = 1$, and $x_4 = 0$. We see that this solution does satisfy the constraints of (41) and hence the optimal solution to (41) is:

$$x_1^* = 0, \; x_2^* = 2, \; x_3^* = 1, \; x_4^* = 9, \text{ and } z^* = 9.$$

If we wished, we could check to see if there are any other optimal solutions by checking the other possibilities. In this problem the truncated value of the linear-programming problem gave the value of z for the optimal hyperplane.

In general, one may have to reduce z° several times. However, the linear-programming solution for the *objective function* is usually close, and frequently the values of x_j^* *are reasonably close.*

Example 2:

$$\begin{aligned} &Max\ z = 10x_1 + 8x_2 + 7x_3, \\ &8x_1 + 7x_2 + 5x_3 \leqslant 34, \\ &x_1 + x_2 + x_3 \leqslant 6, \\ &x_1,\ x_2,\ x_3 \geqslant 0,\ \text{integer}. \end{aligned} \tag{42}$$

The linear programming solution is $x_1 = 1.333$, $x_3 = 4.667$, $z = 46$. Because $c_1 \neq 1$, we convert (42) to the following:

$$\begin{aligned} &Max\ z = x_0 + 10x_1 + 8x_2 + 7x_3, \\ &8x_1 + 7x_2 + 5x_3 \leqslant 34, \\ &x_1 + x_2 + x_3 \leqslant 6, \\ &x_1,\ x_2,\ x_3 \geqslant 0,\ \text{integer}, \\ &x_0 = 0 \end{aligned} \tag{43}$$

The following bounds are easily derived:

$$0 \leqslant x_0 \leqslant 0, \quad 0 \leqslant x_1 \leqslant 4, \quad 0 \leqslant x_2 \leqslant 4, \text{ and } 0 \leqslant x_3 \leqslant 6.$$

We then calculate:

$$\Lambda_0 = 0,\ \Lambda_1 = 40,\ \Lambda_2 = 72,\ \Lambda_3 = 114,\ z^\circ = 46;$$

$$x_3^*(46) = 0,\ 1,\ 2,\ \ldots,\ \min\left[6,\ \frac{46}{7}\right] = 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6.$$

We give a sample calculation for $x_3^* = 0$. If this is the case then $\lambda = 46$ and $x_2^* = \left[\frac{46-41}{8}\right] + 1, \ldots, 4 = 1, 2, 3, 4$. If $x_2^* = 1$, then $\lambda = 38$ and $x_1^* = \left[\frac{38-1}{10}\right] + 1, \ldots, 3 = 4, 3$. Since $4 > 3$, this corresponds to $g_1(38) = -\infty$. If $x_2^* = 2$, then $\lambda = 30$ and $x_1^* = \left[\frac{30-1}{10}\right] + 1, \ldots, 3 = 3$. This gives a potential solution $x_1 = 3$, $x_2 = 2$, $x_3 = 0$. However, this violates one of the constraints of (42). If $x_2^* = 3$, there is no solution, since $g_1(22) = -\infty$, and if $x_2^* = 4$, there is again no solution. Repeating this entire process for the remaining values of x_3^* will also yield similar results.

We next try $z_k = 45$. We again find

$$x_3^*(45) = 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6.$$

Noting that x_3 in the linear-programming solution was between 4 and 5, a good computational strategy is to first try 6 rather than zero. If $x_3^* = 6$, then $\lambda = 3$. This yields $x_2^* = 0$ and x_1^* nonexistent. We next try $x_3^* = 5$. This yields $\lambda = 10$ and $x_2^* = 0, 1$. If $x_2^* = 0$, then $\lambda = 10$ and $x_1^* = \left[\frac{10-1}{10}\right] + 1, \ldots, 1 = 1$. We have a potential solution $x_1^* = 1$, $x_2^* = 0$, and $x_3^* = 5$. This is found to satisfy the constraints of (42) and hence we have found the optimal solution.

6. COMPUTATIONAL RESULTS

In order to gain some insight into the computational feasibility of the method developed in this paper, an experimental code was written to solve problems with randomly generated

data. A program to generate problems such that (1) always has a feasible noninteger solution was developed. The matrix A had full density. This averaged about 90% dense for all matrices generated because of some random generation of zeros. The range of values for a_{ij}, the elements of A, was $0 \leqslant a_{ij} \leqslant 10$ for all the problems generated. For the majority of problems tested, the values of c_j, the objective function coefficients, were in the range $1 \leqslant c_j \leqslant 4$. These were also randomly generated. On the 15×50 problem, coefficients in the range $0 \leqslant c_j \leqslant 50$ were used. However, only ten of the coefficients were positive. The values of the coefficients for b_i, the components of the requirements vector b, were computed as follows. A set of values of a "feasible point", x_f was randomly generated in the range (0, 4). If x_{fj} are the components of x_f, then

$$b_i = \sum_j a_{ij}\, x_{fj} + 1$$

By this means the generation of a problem with a feasible solution was guaranteed.

In the implementation of the algorithm, step 2, the determination of lower and upper bounds, was carried out using equation (5), i.e., a simplex calculation for the larger problems. For smaller problems, a cruder method for bound determination was used. All calculations were done on a CDC CYBER 70, Model 72, which is a moderate speed computer. The results are given in Table 1.

TABLE 1 — *Computational Results*

m	n	No. of Problems	Mean Time (S)	Least Time	Greatest Time	Bounds by Simplex
3	10	10	0.2	0.05	0.7	No
4	10	10	5.7	0.08	28.6	No
5	10	10	22.4	0.08	118.4	No
5	15	10	14.8	1.3	89.8	No
5	20	1	0.5	—	—	No
3	15	5	0.9	0.09	3.0	No
4	15	10	13.2	0.2	50.3	No
4	20	3	32.5	0.1	97.2	No
5	16	1	0.1	—	—	No
5	21	1	1.0	—	—	No
4	24	1	5.6	—	—	No
4	28	1	0.2	—	—	No
5	32	1	0.2	—	—	No
8	10	20	2.18	2.0	3.0	Yes
10	20	25	12.5	5.0	71.0	Yes
10	25	5	140.8	11.0	639.0	Yes
10	30	7	207.1	19.0	1284.0	Yes
12	33	2	26.5	21.0	32	Yes
15	40	4	342.8	54.0	782.0	Yes
15	50	1	218.0	—	—	Yes

The results of Table 1 indicate that the method has been used successfully on problems of relatively moderate size. Further tests on larger problems have not been made for a number of reasons. First, program and storage optimization would significantly alter the times presented in Table 1. More importantly, the wide variation in times found for problems of the same size

is probably due to the nature of the problems, i.e., a matrix A with high density and with randomly generated coefficient entries. Real-world problems would be much more sparse and highly structured. Hence, no definite conclusions can be drawn from Table 1 about how the method would operate on such problems. Unfortunately, such problems are not available to the authors for testing. The results of Table 1 are sufficiently encouraging, however, that this method should be explored further on real problems.

ABSTRACT

An exact method for solving all-integer linear-programming problems is presented. Dynamic-programming methodology is used to search efficiently candidate hyperplanes for the optimal feasible integer solution. The explosive storage requirements for high-dimensional dynamic programming are avoided by the development of an analytic representation of the optimal allocation at each stage. Computational results for problems of small to moderate size are also presented.

REFERENCES

[1] Cooper, L., "Hyperplane Search Algorithms for the Solution of Integer Programming Problems," *IEEE Transactions on Systems, Man and Cybernetics, SMC-3,* 234-240, (1973).

[2] Nemhauser, G. L., *Introduction to Dynamic Programming,* (John Wiley, New York, 1966).

SOLUTION OF CONTINUOUS-TIME MARKOVIAN DECISION MODELS USING INFINITE LINEAR PROGRAMMMING

Prasadarao Kakumanu

School of Management
University of Scranton
Scranton, Pennsylvania

ABSTRACT

Infinite-horizon, countable-state, continuous-time Markovian decision models are solved by formulating as a pair of infinite linear-programming problems. Expected discounted and average returns are considered as criterion functions. For both criterion functions, the existence of deterministic optimal stationary policies is established by solving the associated infinite linear-programming problems. Computational procedures for finite state and action sets are discussed by considering associated finite linear-programming problems.

1. INTRODUCTION

In this paper, we develop a methodology that can be used to analyze stochastic dynamic systems, such as health, transportation, educational, economic, and production systems. These systems change continuously, due to changes in the environment in which they operate. Some of the factors that influence these systems are controllable and others are not, hence the behavior of the systems cannot be predicted completely. This makes it appropriate to treat the systems as stochastic rather than deterministic. Since some of the factors that influence a system are controllable, our object is to determine how to adjust these controllable factors, so that the system operates satisfactorily. We assume that the possible states of the system are countable and that the available actions are finite; they are denoted, respectively by S and A. The changes in the system are governed by a set of known numbers $q_{ij}(a)$ $(i, j \in S, a \in A)$, called the transition rates. The transition rates may be interpreted as probabilities; that is, if at time t the system is in state i, $i \in S$, then the approximate probability that the system is in state $j \neq i$ after time δt is given by $q_{ij}(a)\delta t$, and the probability that it remains in the same state is $[1 - \Sigma_{j \neq i} q_{ij}(a)\delta t]$. The system may be observed at any time and classified into one of the possible states i, $i \in S$. If the system is in state $i \in S$ and an action $a \in A$ is taken, then the system receives a bounded return $r(i, a)$ per unit time and moves to the new state $j \in S$ governed by the transition rates $q_{ij}(a)$. Under certain general conditions on $\{q_{ij}(a)\}$, the trajectory of the stochastic dynamic system can be represented by a Markov process. In view of this, we use the words system and process synonymously.

We are concerned with the optimal control of the type of stochastic systems described above. For controlling such systems we need a rule that prescribes the action to be taken; this rule, denoted by π, is called a policy. It is assumed that the policy π is specified by a family of measurable functions $\{d_{ia}(t)\}$, where, for each $i \in S$

$$\sum_a d_{ia}(\cdot) = 1, \text{ and } d_{ia}(\cdot) \geq 0, a \in A. \tag{1.1}$$

Thus the function $d_{ia}(t)$ may be interpreted as the probability of taking an action a, $a \in A$, at time t, given that the current state of the system is $i \in S$. A policy π is called Markovian if $d_{ia}(t) = 1$ or 0, and it is called stationary if $d_{ia}(\cdot)$ is independent of t and satisfies (1.1). A stationary policy is called a deterministic stationary policy if, for each i, $d_{ia} = 1$ or 0, $i \in S$, $a \in A$. That is, a deterministic stationary policy specifies a single action depending upon the state of the processes but independent of time t.

When a policy π is applied, the transition rate from a state i to a state j is given by

$$q_{ij}(t, \pi) = \sum_a q_{ij}(a)\, d_{ia}(t), \quad i, j \in S. \tag{1.2}$$

Let $Q(t, \pi)$ be the transition rate matrix whose $(i,j)^{th}$ element is given by (1.2). For each $i \in S$, the transition rates are assumed to satisfy:

$$\sum_j q_{ij}(t, \pi) = 0, \quad q_{ij}(t, \pi) \geqslant 0, \quad j \neq i, \text{ and} \tag{1.3}$$

$$0 < | q_{ii}(t, \pi) | \leqslant M < \infty. \tag{1.4}$$

Under these assumptions, it was shown in Ref. [13] that, for any given Markov policy π, a unique transition probability matrix

$$F(s,t,\pi) = \{f_{ij}(s,t,\pi), \ i, j \in S\} \tag{1.5}$$

exists, and that for almost all $t \geqslant s$ it satisfies the Kolmogorov differential equations:

$$\frac{\partial F(s, t, \pi)}{\partial t} = F(s,t,\pi)\, Q(t,\pi), \text{ with } F(s, s, \pi) = I. \tag{1.6}$$

If π is a stationary policy, then the transition rate matrix Q is independent of t, and the transition probability matrix F defines a time-homogeneous Markov process $x(t, \pi)$.

The expected rate of return out of a state i, when the Markov policy π is given by

$$r(i, t, \pi) = \sum_a r(i, a) d_{ia}(t), \quad t \geqslant 0, \quad i \in S. \tag{1.7}$$

Since the rate of return $r(i,a)$ is uniformly bounded, it is obvious from (1.7) that $r(i, t, \pi)$ is uniformly bounded in i and π. For any Markov policy π, using the above definitions and notations, we define two economic criterion functions. The first one, the total expected discounted-return function, which is given by:

$$\Psi(i, \alpha, \pi) = \int_0^\infty e^{-\alpha t} \sum_j f_{ij}(o, t, \pi)\, r(j, t, \pi)\, dt, \quad i \in S, \tag{1.8}$$

where $\alpha > 0$, is called the discount factor. The other, the long-run expected average-return function, is defined by:

$$\Phi(i, \pi) = \lim_{T \to \infty} T^{-1} \int_0^T \sum_j f_{ij}(o, t, \pi)\, r(j, t, \pi)\, dt, \quad i \in S. \tag{1.9}$$

If the limit in (1.9) does not exist for any $i \in S$, then we set $\Phi(i, \pi) = -\infty$ for that i.

A policy π^* is called an *α-discounted optimal policy* if $\psi(i, \alpha, \pi^*) \geqslant \psi(i, \alpha, \pi)$, $i \in S$, and it is called an *average optimal policy* if $\Phi(i, \pi^*) \geqslant \Phi(i, \pi)$, $i \in S$, where π is any Markov policy.

The object of this paper is to show the existence of an optimal deterministic stationary policy for the discounted and the average-return case, by considering a separate pair of infinite

linear-programming problems. A linear-programming problem is called an infinite linear-programming problem if the number of variables and constraints are countably infinite [6].

The methodology developed here is important mainly for two cases which are currently under investigation by the author. In the first case, using the results developed by Fox [9], and Gustafson and Kortanek [10], we can obtain an approximate optimal policy, along with the corresponding return for the criterion functions (1.8) and (1.9) by considering a series of finite linear-programming problems. It is also possible to estimate the error between the approximate and optimal solutions, and the convergence rate. In the second case, that of the discounted model, the methodology can be used to parametrize the discount factor. This enables us to identify the interval in which an optimal policy is invariant. The parametric analysis of the discount factor will be, in a certain sense, an extension of the results given by Denardo [3], Howard (Ref. [11], p. 88), Miller and Veinott [16], Mine and Osaki (Ref. [18], p. 22), and Veinott [22].

Semi-Markovian or Markov renewal decision processes are also used to study stochastic systems (see Ross, Ref. [20], p. 156). In this case, the decisions are made only at the instant of transitions, whereas in the continuous-time Markovian decisions models the decision are made at any time. As pointed out by Doshi [5], the optimal policy we obtain using semi-Markovian decisions might result in suboptimal operation of the system. In addition to this, with continuous-time Markovian decision models we need less data than we do with semi-Markovian decision models to analyze a stochastic system. For these reasons, continuous-time Markovian decision models are preferable to semi-Markovian decision models for studying certain stochastic systems.

In Section 2, the existence of α-discounted optimal deterministic stationary policies is proved for all Markov policies. In Section 3, the analogous results for the average-return criterion function is established. Finally, in Section 4, the computational aspects of the results obtained in the earlier sections are discussed when the state space is finite.

2. COUNTABLE-STATE DISCOUNTED-RETURN MODEL

In this section, the following results are established by formulating the expected discounted-return model as a pair of infinite linear programs called primal and dual problems. Primal and dual problems always have bounded feasible solutions. There is one-to-one correspondence between the set of dual feasible solutions and the set of stationary policies. Both problems have bounded optimal feasible solutions such that the corresponding values of their objective functions are equal. At least one dual optimal feasible solution is capable of interpretation as an α-discounted optimal deterministic stationary policy. Finally, a sufficient condition is given which will guarantee that every dual optimal feasible solution will yield an α-discounted optimal deterministic stationary policy.

The methodology given in this section provides the means of obtaining an approximate optimal policy for the criterion function (1.8), and of conducting parametric analysis of the discount factor α. These could not have been done by using the results given in Ref. [14] and the finite state policy interaction methods because of computational difficulty. The results obtained here are a generalization of discrete-time results proved by Evans [8]. Some techniques given by him will be used to establish our results.

Let $\{\gamma_j;\ j \in S\}$ be a set of positive number such that $\sum_j \gamma_j = 1$, and g be a vector defined on S with bounded components. Consider the following pair of infinite programs:

PRIMAL I	DUAL I
$\text{Inf} \sum_j \gamma_j g_j$	$\sup \sum_i \sum_a w_{ia}\, r(i, a)$
subject to	subject to
$\alpha g_i - \sum_j q_{ij}(a)\, g_j \geqslant r(i,a)$,	$\sum_i \sum_a w_{ia} [\alpha \delta_{ij} = q_{ij}(a)] = \gamma_j$,
$i \in S,\ a \in A.$	$w_{ja} \geqslant 0,\ j \in S,\ a \in A.$

LEMMA 2.1 The primal always has a bounded feasible solution.

PROOF: Let $\psi(i, \alpha \pi^*) = \sup_\pi \psi(i, \alpha, \pi)$, $i \in S$. Then, it is known [14] that $\psi(i, \alpha, \pi^*)$, $i \in S$, is the unique solution of the functional equation

$$\alpha g_i = \max_a \left[r(i,a) + \sum_j q_{ij}(a)\, g_j \right], \qquad i \in S.$$

Taking $g_i = \psi(i, \alpha, \pi^*)$, $i \in S$, satisfies the primal constraints. Hence, $\{\psi(i, \alpha, \pi^*)\}$ is a primal feasible solution. The value of the objective function is given by

$$\sum_i \gamma_i |\psi(i, \alpha, \pi^*)| = \sum_j \gamma_i \left| \int_0^\infty e^{-\alpha t} \sum_j f_{ij}(t, \pi^*)\, r(j, \pi^*)\, dt \right|$$

$$\leqslant \left(\sum_i \gamma_i \right) \frac{1}{\alpha} \left[\sup_j |r(j, \pi^*)| \right]$$

Since $\sum_i \gamma_i = 1$, the $r(j, \pi^*)$ are uniformly bounded, and $\alpha > 0$, $\sum_i \gamma_i \psi(i, \alpha, \pi^*)$ is a finite number.

LEMMA 2.2 The dual has a bounded feasible solution.

PROOF: Let π be any stationary policy specified by $\{d_{ia}\}$. Define

$$u_{ia} = \sum_l \gamma_l \int_0^\infty e^{-\alpha t} f_{li}(t, \pi)\, d_{ia}\, dt, \quad i \in S,\ a \in A. \tag{2.1}$$

Now we will show that $\{u_{ia}\}$is a bounded dual feasible solution. It is obvious that $u_{ia} \geqslant 0$, $i \in S$, $a \in A$. Substituting u_{ia} from (2.1) for w_{ia} in the left hand side of the dual constraint, we arrive at:

$$\sum_i \sum_a [\alpha \delta_{ij} - q_{ij}(a)] \sum_l \gamma_l \int_0^\infty e^{-\alpha t} f_{li}(t, \pi)\, d_{ia}\, dt, \quad j \in S.$$

Since the series converges absolutely, we obtain after some simplification

$$\sum_l \gamma_l \int_0^\infty e^{-\alpha t} \left[\alpha f_{lj}(t, \pi) - \sum_i f_{li}(t, \pi)\, q_{ij}(\pi) \right] dt, \quad j \in S.$$

Using (1.6) and integrating by parts, we can show that this is equivalent to γ_j, $j \in S$. Substituting the value of $\{u_{ia}\}$ from (2.1) in the dual objective function for w_{ia}, we obtain after some simplification

$$\sum_i \sum_a u_{ia}\, r(i, a) = \sum_l \gamma_l \psi(l, \alpha, \pi). \tag{2.2}$$

Since $\sum_l \gamma_l = 1$, and $\psi(l, \alpha, \pi)$ is uniformly bounded in l, α, and π, the righthand side is finite. Hence, $\{u_{ia}\}$ is a bounded dual feasible solution.

THEOREM 2.1: There is one-to-one correspondence between the set of dual feasible solutions and the set of stationary policies.

PROOF: Let $\{w_{ja}\}$ be any dual feasible solution. First we will show $\sum_a w_{ja} > 0$, $j \in S$. The dual constraints may be written as

$$\left(\sum_a w_{ja}\right)[\alpha - q_{jj}(a)] - \sum_{i \neq j} \sum_a w_{ia}\, q_{ij}(a) = \gamma_j, \quad j \in S. \tag{2.3}$$

Since $\gamma_j > 0$, $j \in S$, and $\{q_{ij}(a)\}$ satisfies (1.3), we have

$$\sum_a w_{ja} > 0. \tag{2.4}$$

Define

$$d_{ia} = \frac{w_{ia}}{\sum_b w_{ib}}, \qquad i \in S,\ a \in A. \tag{2.5}$$

By definition, the set $\{d_{ia}\}$ defines a stationary policy, say δ. If $\{u_{ia}\}$ is the set of values given by (2.1) when the policy δ is applied, as shown in Lemma 2.2, $\{u_{ia}\}$ is a dual feasible solution. If

$$u_{ia} = w_{ia}, \qquad i \in S,\ a \in A, \tag{2.6}$$

then we can conclude that there is one-to-one correspondence between the set of feasible solutions and the set of stationary policies. Now we will show that (2.6) is true.

Let $w_i = \sum_a w_{ia}$, $i \in S$, and W, and γ represent the row vectors whose i^{th} elements are given by w_i and γ_i, $i \in S$, respectively. Using this notation and (1.2), we may write 2.3 as

$$W[\alpha I - Q(\delta)] = \gamma,$$

since $[\alpha I - Q(\delta)]$ is the resolvent of the operator $Q(\delta)$. From Dynkin (Ref. [7], p. 24) we obtain

$$W = \int_0^\infty e^{-\alpha t} \gamma F(t, \delta)\, dt.$$

The i^{th} component of W is given by

$$\sum_a w_{ia} = w_i = \sum_l \gamma_l \int_0^\infty e^{-\alpha t} f_{li}(t, \delta)\, dt, \quad i \in S.$$

Using (2.1) and (2.5) we obtain:

$$\sum_a w_{ia} = \sum_a u_{ia}, \qquad i \in S.$$

From this, (2.1), and (2.3) we conclude that (2.6) is true.

LEMMA 2.3: If $\{g_l\}$ and $\{w_{ia}\}$ are, respectively, primal and dual feasible solutions, then

$$\sum_l \gamma_l g_l \geq \sum_i \sum_a w_{ia}\, r(i, a). \tag{2.7}$$

Since the infinite series that appear in both primal and dual are absolutely convergent, the proof of the lemma is similar to the corresponding result in finite linear programming [21].

Since the primal and dual problems have bounded feasible solutions, the optimal solutions for both problems will always exist. In the following theorem, we establish that there is no duality gap, that is that the values of the objective functions for the corresponding optimal feasible solutions are equal. Then, using the results developed here, we establish the existence of an α-discounted optimal deterministic stationary policy.

THEOREM 2.2: There is no duality gap.

PROOF: Let $\{w_{ia}^*\}$ be a dual optimal feasible solution and π^* be the corresponding stationary policy obtained from $\{w_{ia}\}$ through (2.5). Now, following the argument given in Lemma 2.2, we obtain

$$\sum_i \sum_a w_{ia}^* r(i, a) = \sum \gamma_i \psi(i, \alpha, \pi^*). \tag{2.8}$$

Since $\gamma_l > 0$, $l \in S$, and the same π^* maximizes $\psi(l, \alpha, \pi)$ over all stationary policies for every $l \in S$, we obtain

$$\psi(i, \alpha, \pi^*) = \sup_{\pi} \{\psi(i, \alpha, \pi)\}, \quad i \in S.$$

In Lemma 2.1, it was shown that $g_i^* = \psi(i, \alpha \pi^*)$, $i \in S$, is a primal feasible solution. From (2.8) we conclude that $\{g_i^*\}$ is a primal optimal feasible solution, and that the values of the primal and dual objective functions are equal.

THEOREM 2.3: An α-discounted optimal deterministic stationary policy always exists, and there is at least one dual optimal solution that will give this policy.

PROOF: Let $\{g_i^*\}$ be the optimal primal feasible solution. Consider the sets

$$T_l = \left\{ a_l \middle| \alpha g_l^* = r(l, a_l) + \sum_j q_{lj}(a_l)\, g_j^*,\ a_l \in A \right\}, \quad l \in S,$$

where $a_l \in A$ is the action taken when the state of the processes is $l \in S$. For each state $l \in S$ it can be easily shown that T_l is not an empty set [14]. Let us define

$$a_i^* = \min_l \{l | a_l \in T_i\},$$

$$v_{ia}^* = \begin{cases} 1 & \text{if } a = a_i^*, \text{and} \\ 0 & \text{otherwise.} \end{cases}$$

Then a policy σ^* can be obtained from $\{v_{ia}^*\}$ using (2.5), and it is clear that σ^* is a deterministic stationary policy. As in Theorem 2.1, we can show that $\{v_{ia}^*\}$ is a dual feasible solution. Using (2.2) and Theorem 2.2, we may write the dual objective function as

$$\sum_i \sum_a u_{ia}^* r(i, a) = \sum \gamma_i \psi(i, \alpha, \sigma^*) = \sum_i \gamma_i g_i^*.$$

Hence, σ^* is an optimal deterministic stationary policy, and $\{v_{ia}^*\}$ is the optimal dual feasible solution that will yield this policy.

In Theorem 2.6, we give a sufficient condition that will guarantee the optimal deterministic stationary policy for every dual optimal feasible solution. We can prove the following two

theorems easily using Lemma 2.3, Theorem 2.2 and the duality theory of finite linear programming.

THEOREM 2.4: The necessary and sufficient condition that any primal $\{g_i\}$ and dual $\{w_{ia}\}$ feasible solutions are optimal to respective problems is that

$$\sum_i \gamma g_i = \sum_i \sum_a w_{ia}\, r(i,\, a).$$

THEOREM 2.5: Any pair of primal $\{g_i\}$ and dual $\{w_{ia}\}$ feasible solutions is optimal to respective problems if and only if

$$w_{ia}\left[\alpha g_i - \sum_j q_{ij}(a) g_j - r(i,a)\right] = 0,\ i \in S,\ a \in A,$$

and

$$g_i\left[\sum_i \sum_a w_{ia}\, \alpha \delta_{ij} - q_{ij}(a)\right] = 0,\ i \in S.$$

THEOREM 2.6: If there is a primal optimal solution $\{g_i^*\}$ such that

$$\alpha g_i^* = r(i,\, a) + \sum_j q_{ij}(a) g_j^* \tag{2.9}$$

for only one $a \in A$ for each $i \in S$, then every dual optimal solution yields an α-discounted optimal deterministic stationary policy.

PROOF: Let $a_i^* \in A$ be the unique a for which (2.9) holds, that is

$$\alpha g^* = r(i,\, a_i^*) + \sum_j q_{ij}(a_i^*) g_j^*,\ i \in S. \tag{2.10}$$

For any other $a \in A$, we have

$$\alpha g_i^* > r(i,\, a) + \sum_j q_{ij}(a)\, g_j^*,\ i \in S. \tag{2.11}$$

Let $\{w^*_{ia}\}$ be any dual optimal feasible solution. Then from (2.4) we have

$$\sum_a w_{ia}^* > 0,\ i \in S. \tag{2.12}$$

From Theorem 2.5, (2.10), and (2.11), for each $i \in S$, we have

(2.13) $w_{ia}^* \geqslant 0$ if $a = a^*$, and

$= 0$ for all other $a \in A$.

From (2.12) and (2.13), we obtain for each $i \in S$

$w_{ia}^* > 0$ if $a = a^*$, and

$= 0$ for all other $a \in A$.

Now define

$$d_{ia}^{\bullet} = \frac{w_{ia}^{\bullet}}{\sum_b w_{ib}^{\bullet}} = 1 \text{ or } 0, \quad i \in S, \ a \in A.$$

Hence, by Theorem 2.1, σ^* is an α-discounted optimal deterministic stationary policy and the corresponding optimal return is give by $g_i^{\bullet} = \psi(i, \alpha, \sigma^*)$, $i \in S$.

3. COUNTABLE-STATE AVERAGE-RETURN MODEL

In this section, by considering a pair of infinite linear programs we establish the existence of an average optimal deterministic stationary policy. The methodology that will be developed in this section is analogous to that given in Section 2. In view of this, we will briefly indicate the proof for the main results and the rest will be stated without proofs. Consider a pair of infinite linear programs associated with the average-return model:

PRIMAL II	DUAL II
Inf h	Sup $\sum_i \sum_a x_{ia}\, r(i,a)$
subject to	subject to
$h - \sum_j q_{ij}(a)\, v_j \geq r(i,a)$, $i \in S, \ a \in A$.	$\sum_i \sum_a x_{ia}\, q_{ij}(a) = 0, \ j \in S$, and $\sum_i \sum_a x_{ia} = 1$, $x_{ia} \geq 0, \ i \in S, \ a \in A$.

Throughout this section we need the following assumption:

ASSUMPTION 3.1: For each deterministic stationary policy π, the resulting Markov process $x(t, \pi)$ is positive recurrent with only one recurrent class.

THEOREM 3.1: Primal and dual problems have bounded feasible solutions.

PROOF: The existence of bounded primal feasible solutions is proved by use of the argument given in Lemma 2.1, and Theorem 3.1 of Ref. [15]. The existence of dual feasible solutions can be shown as follow. Let π be any stationary policy specified by the set $\{d_{ia}\}$. Consider

$$y_{ia} = \lim_{T \to \infty} \frac{1}{T} \int_0^T f_{li}(t, \pi)\, d_{ia}\, dt, \quad l \in S. \tag{3.1}$$

It is obvious that $\{y_{ia}\}$ satisfies nonnegative restrictions and that $\sum_i \sum_a y_{ia} = 1$. Substituting the value of y_{ia} from (3.1) in the left-hand side of the first dual constraint, after some simplification we arrive at

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \left[\sum_i f_{li}(t, \pi) q_{ij}(\pi) \right] dt, \quad j \in S.$$

We can show that this is equal to zero by using (1.6) and integrating by parts. Since $r(i, a)$ is uniformly bounded in i and a, it follows that $\sum_i \sum_a y_{ia}\, r(i, a)$ is finite. Hence, $\{y_{ia}\}$ is a bounded dual feasible solution.

THEOREM 3.2: There is one-to-one correspondence between the set of dual feasible solutions and the set of stationary policies.

PROOF: Let $\{x_{ia}\}$ be any dual feasible solution. As in Theorem 2.1, we can show that

$$d_{ia} = \frac{x_{ia}}{\sum_b x_{ib}}, \qquad i \in S,\ a \in A, \tag{3.2}$$

defines the probability distribution on A for each $i \in S$. This shows that the dual feasible solution $\{x_{ia}\}$ yields a stationary policy, say π.

Let $\rho_i(\pi)$, $i \in S$, be the steady-state probabilities of processes when the stationary policy π is used. Under Assumption 3.1 $\rho_i(\pi), i \in S$, satisfies the following conditions:

$$\sum_i \rho_i(\pi)\ q_{ij}(\pi) = 0,\ i \in S, \tag{3.3}$$

$$\sum_i \rho_i(\pi) = 1, \text{ and } \rho_i(\pi) > 0,\ i \in S.$$

Using (1.2), (3.2), (3.3), and dual feasible solution $\{x_{ia}\}$, we obtain

$$\sum_a x_{ia} = \rho_i(\pi),\ i \in S. \tag{3.4}$$

Noting that $\rho_i(\pi) = \lim_{T\to\infty} T^{-1} \int_0^T f_{li}(t, \pi)\, dt$, $i \in S$, and from (3.1) and (3.4), we have

$$\sum_a x_{ia} = \sum_a y_{ia}, \qquad i \in S. \tag{3.5}$$

From (3.2) we have, for $i \in S$ and $a \in A$, $y_{ia} = \sum_b y_{ib}\, d_{ia} = \sum_b x_{ib}\, d_{ia} = x_{ia}$.

This shows that the stationary policy π yields a set of numbers $\{y_{ia}\}$ from (3.1), which are equal to the dual feasible solution that is used to obtain π, which proves the theorem.

For the primal and dual problems considered in this section, the results corresponding to those of Lemma 2.3 and Theorems 2.4 and 2.5 can be proved by use of the arguments given in Ref. [21] for finite linear programs. In the following theorem, we establish the existence of optimal primal and dual feasible solutions with no duality gap and the existence of an average optimal deterministic stationary policy. The results correspond to Theorems 2.2 and 2.3.

THEOREM 3.3: Both the primal and dual problems have bounded optimal feasible solutions with

$$\text{Inf } h = \sup \sum_i \sum_a x_{ia} r(i, a), \tag{3.6}$$

and there exist's an average optimal deterministic stationary policy.

PROOF: It was shown in Theorem 3.1 that a bounded primal feasible solution (h^*, v_j^*) exists. Let

$$T_l = \left\{ l \middle| h^* = r(l, a_l) + \sum_i q_{li}(a_l)\, v_i^*,\ l \in S \right\},$$

where a_l is the action taken when the process is in state $l \in S$ and at least one such a_l exists. Let $a_i^* = \min\{a_l \mid l \in T_i\}$, $i \in S$, and σ^* be a policy that prescribes the action a_i^* when the process is in state $i \in S$. It is clear that σ^* is a deterministic stationary policy and satisfies.

$$h^* = r(i, \sigma) + \sum_j q_{ij}(\sigma^*) v_j^*, \quad i \in S.$$

Then, from Theorem 3.3 of Ref. [15], we have

$$h^* = \Phi(i, \sigma^*), \quad i \in S. \tag{3.7}$$

Let $\{y_{ia}^*\}$ be the solution obtained from (3.1) when the policy σ^* is used. In Theorem 3.1 it was shown that $\{y_{ia}^*\}$ is a dual bounded feasible solution. The corresponding dual objective function is

$$\sum_i \sum_a y_{ia}^* r(i, a).$$

Substituting the value for y_{ia}^* from (3.1),

$$\sum_i \sum_a \left[\lim_{T \to \infty} T^{-1} \int_0^T f_{li}(t, \pi) d_{ia}^* \, dt \right] r(i, a), \quad l \in S,$$

where $\{d_{ia}^*\}$ corresponds to σ^*. Since all the terms are positive, the limit and the summation signs can be interchanged. Using (1.7) and (1.9), after some simplification we arrive at

$$\sum_i \sum_a y_{ia}^* r(i, a) = \Phi(l, \sigma^*), \quad l \in S. \tag{3.8}$$

We can show from (3.7) and (3.8) that $\{h^*, v_j^*\}$ is a primal optimal solution and $\{y_{ia}^*\}$ is a dual optimal feasible solution as the solutions attain their lower and upper bounds respectively. The policy σ^* is an average optimal deterministic stationary policy.

Under certain conditions similar to those given in Theorem 2.6, it can be proved that every dual optimal feasible solution is capable of interpretation as an average optimal deterministic stationary policy. All the results established in this section can be shown easily when the time parameter t is discrete instead of continuous.

4. FINITE-STATE: CASE COMPUTATIONAL PROCEDURES

In this section we assume that the state and action sets are finite; they are denoted respectively by $S = \{1, 2, \ldots, m\}$ and $A = \{1, 2, \ldots, L\}$, where m and L are finite positive integers. The summations over i, j, and k vary from 1 to m, and over a and b from 1 to L. When the state and action sets are finite, the primal and dual problems given in Sections 2 and 3 can be written as standard linear programs.

The main purpose of discussing the finite state models is to show their use in obtaining the solutions for the infinite linear-programming problems discussed in Sections 2 and 3. This will be done by solving a series of finite-state problems whose solution is known to converge to that of the countably infinite state problem [9]. For a finite state space, Denardo [2], Howard [12], Mine and Osaki [18], and Osaki and Mine [19] have formulated Markovian renewal program problems as finite linear-programming problems. For the reasons stated in Section 1, sometimes it is preferable to study the decision process as a continuous-time Markovian process rather than as the Markov renewal process.

We first discuss the discounted-return model. Mine and Tabata [17] formulated finite-state continuous-time Markovian decision models as standard linear-programming problems by

using various transformations. These transformations limit the use of linear-programming techniques for solving the continuous-time Markovian decision models. The methodology developed in Section 2 will allow us to formulate the continuous-time models as a pair of standard linear programs without the help of any type of transformation. This type of formulation is very important in obtaining the solution to continuous-time Markovian decision models by solving the associated linear programs:

PRIMAL III	DUAL III
$\min \sum_j \gamma_j g_j$	$\max \sum_i \sum_a r(i, a) w_{ia}$
subject to	subject to
$\alpha g_i - \sum_j q_{ij}(a) g_j \geq r(i,a),$	$\sum_a \alpha w_{ja} - \sum_i \sum_a w_{ia}\, q_{ij}(a) = \gamma_j,\ j \in S,$
$i \in S,\ a \in A.$	$w_{ja} \geq 0,\ j \in S,\ a \in A.$

The numbers γ_j, $j \in S$, are strictly postive and add up to unity. The results that were proved in Section 2 hold in particular when the state space is finite. We now show that every dual optimal solution obtained by the simplex method may be interpreted as an α-discounted optimal deterministic stationary policy, and that the corresponding primal optimal solution yields an optimal discounted return.

THEOREM 1.1: There is one-to-one correspondence between the set of dual basic feasible solutions and the set of deterministic stationary policies.

PROOF: Let $\{w_{ia}\}$ be any dual basic feasible solution, that is, at most m of the w_{ia}'s are strictly positive. In Lemma 2.2 we showed that $\sum_a w_{ia} > 0$ for each $i \in S$. Hence, there is only one $w_{ia} > 0$ for each $i \in S$. Let the probability of making action $a \in A$ when the process is in state $i \in S$ be defined by

$$d_{ia} = \frac{w_{ia}}{\sum_b w_{ib}} = 1 \text{ or } 0, \tag{5.1}$$

depending upon whether $w_{ia} > 0$ or not. Then $\{d_{ia}\}$ defines a deterministic stationary policy denoted by σ^*.

Using the argument given in Theorem 2.1, it can be shown that the deterministic stationary policy σ^* corresponds only to $\{w_{ia}\}$. This proves the theorem.

From this theorem, it follows that the dual optimal feasible solution obtained by using the simplex algorithm may be interpreted as an α-discounted optimal deterministic stationary policy. The corresponding optimal expected discounted return is obtained from the primal optimal feasible solution.

We now discuss the finite-state and action-set Markovian decision models and the associated linear programs for the average-return case. As in the discounted case, the infinite linear programs given in Section 3 can be written as regular linear programs given by

PRIMAL IV	DUAL IV
$\min h$	$\max \sum_i \sum_a x_{ia}\, r(i, a)$
subject to	subject to
	$\sum_i \sum_a x_{ia}\, q_{ij}(a) = 0\ j \in S,$ and
$h - \sum_j q_{ij}(a)\, v_j \geq r(i,a),$	$\sum_i \sum_a x_{ia} = 1,$
$i \in S,\ a \in A.$	$x_{ia} \geq 0,\ i \in S,\ a \in A.$

It should be noted that the dual problem has $(m + 1)$ constraints, but it is easy to show that at least one of the dual constraints is a redundant constraint. Hence any basic feasible solution can have at most m of the x_{ia} different from zero. Since any dual feasible solution $\{x_{ia}\}$ must satisfy $\sum_a x_{ia} > 0$ for each $i \in S$, we conclude that all basic feasible solutions are non-degenerate. It can be shown that each dual basic feasible solution corresponds to a deterministic stationary policy and vice versa. Every dual basic feasible solution can be interpreted as a deterministic stationary policy by defining

$$d_{ia} = \frac{x_{ia}}{\sum_b x_{ib}}, \quad i \in S, \ a \in A.$$

If the simplex method is used to solve the dual problem, we will always get an optimal basic feasible solution. This optimal solution yields an average optimal deterministic stationary policy, and the corresponding primal optimal solution will give the average optimal return $h^* = \Phi(i, \sigma^*)$, $i \in S$.

When the state and action sets are finite, the above methodology will give an alternate procedure for finding the optimal policies and the corresponding returns. The methodology also allows us to conduct a sensitivity analysis of the various parameters in the model, using existing linear-programming codes.

BIBLIOGRAPHY

[1] Blackwell, D., "Discounted Dynamic Programming, "Annals of Mathematical Statistics, *36*, 226-235 (1965).

[2] Denardo, E. V., and B. L. Fox, "Multichain Markov Renewal Programs," SIAM Journal on Applied Mathematics, *16*, 468-487 (1968).

[3] Denardo, E. V., "Markov Renewal Program with Small Interest Rates," Annals of Mathematical Statistics, *42*, 477-496 (1971).

[4] Derman, C., *Finite State Markovian Decision Processes* (Academic Press, New York, 1970).

[5] Doshi, B., "On the Continuous-time Control of Queues," Technical Report, Department of Statistics, Rutgers University, New Brunswick, N. J. (1975).

[6] Duffin, R. J., and L. A. Karlovitz, "An Infinite Program with a Duality Gap," Management Science, *12*, 122-134 (1965).

[7] Dynkin, E. B. *Markov Processes - I* (Academic Press, New York, 1965).

[8] Evans, J. P., "Duality in Markov Decision Problems with Countable Action and State Space," Management Science, *15*, 626-638 (1968).

[9] Fox, B. L., "Finite-State Approximations to Denumerable-State Dynamic Programming," Journal of Mathematical Analysis and Applications, 665-670 (1971).

[10] Gustafson, S. A., and K. O. Kortanek, "Numerical Treatment of a Class of Semi-Infinite Programming Problems," Naval Research Logistics Quarterly, *20*, 477-504 (1973).

[11] Howard, R. A., *Dynamic Programming and Markov Processes* (The M.I.T. Press, Cambridge, Mass., 1960).

[12] Howard, R. A., "Research in Semi-Markovian Decision Structures," Journal of the Operations Research Society of Japan, *6*, 163-199 (1964).

[13] Kakumanu, P., "Continuous Time Markov Decision Models with Applications to Optimization Problems," Technical Report 63, Cornell University, Ithaca, New York (1969).

[14] Kakumanu, P., "Continuously Discounted Markov Decision Model with Countable State and Action Space," Annals of Mathematical Statistics, *42*, 919-926 (1971).

[15] Kakumanu, P., "Continuous Time Recurrent Markovian Decision Processes Average Return Criterion," Journal of Mathematical Analysis and Applications, *52*, 173-188 (1975).

[16] Miller, B. L., and A. F. Veinott, Jr., "Discrete Dynamic Programming with a Small Interest Rate," Annals of Mathematical Statistics, 366-370 (1969).

[17] Mine, H., and Y. Tabata, "Linear Programming and Continuous Markovian Decision Problems," Journal of Applied Probability, *7*, 657-666 (1970).

[18] Mine, H., and S. Osaki, *Markovian Decision Processes*, (American Elsevier Publishing Company, New York, 1970).

[19] Osaki, S., and H. Mine, "Linear Programming Algorithm for Semi-Markovian Decision Processes," Journal of Mathematical Analysis and Applications, *22*, 356-381 (1968).

[20] Ross, S. M., *Applied Probability Models with Optimization Applications*, (Holden-Day, San Francisco, CA, 1970).

[21] Simonnard, M., *Linear Programming* (Prentice-Hall, Englewood Cliffs, NJ, 1966).

[22] Veinott, A. F., Jr., "Discrete Dynamic Programming with Sensitive Discount Optimality Criteria," Annals of Mathematical Statistics, *40*, 1635-1660 (1969).

RENEWAL PROCESSES OF PHASE TYPE

Marcel F. Neuts

University of Delaware
Newark, Delaware

ABSTRACT

This paper discusses a class of analytically and numerically tractable renewal processes, which generalize the Poisson process. When used to describe interarrival or service times in queues, these renewal processes lead to computationally explicit solutions which involve only real arithmetic. Previous modifications of the Poisson process, based on the Erlang or the hyperexponential distributions, appear as particular cases.

1. INTRODUCTION

In many stochastic models, tractable analytic or numerical results are usually only obtained if certain random variables are assumed to have a negative exponential distribution. This accounts for the wide use of the Poisson process as an arrival process in the analysis of queues and counters, the birth-and-death assumptions underlying epidemic models, and the negative exponential durations ascribed to service times, lead times, headway in traffic, and a large variety of other random time intervals. The classical memory-less property, which eliminates the drastic growth in dimensionality due to conditioning, is the underying source of all simplifications that we owe to the negative exponential distribution.

The limitations of the exponential distribution in modeling real durations are well-known. A large probability is assigned to the shorter time intervals, and the proper unimodality or multimodality of many real situations cannot be represented. This was recognized by A. K. Erlang. It led him to introduce the probability distributions which bear his name. In practice, it is now common to assume Erlang or hyperexponential (finite mixtures of negative exponentials) distributions to model random time durations which are too far removed from the exponential case. These distributions have a greater versatility and allow for relatively tractable expressions under repeated conditioning.

With these desirable properties in mind, the author [3] introduced the *probability distributions of phase type*, of which the Erlang and hyperexponential distributions are very special cases. Discusssions of the algorithmic simplifications introduced in the study of some problems in queues and branching processes are given in Refs. [4], [5], [6] and [7].

The present paper deals specifically with renewal processes in which the distribution of the time between renewals is of phase type. The material developed here is basic in the discussion

*This research was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-72-2350 B. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copy right notation hereon.

of certain queueing models which, to date, have not been solved in a computationally convenient form [7].

HISTORICAL BACKGROUND

We first review some definitions and state some results established earlier. In this paper we only concern ourselves with probability distributions of phase type on $[0, \infty)$, bearing in mind that there is an entirely analogous development of probability distributions of phase type on the nonnegative integers.

We consider an $(m+1)$-state, continuous-parameter Markov chain with states $1, \ldots, m+1$, whose infinitesimal generator Q has the form

$$Q = \begin{vmatrix} T & \mathbf{T}^0 \\ \mathbf{O} & O \end{vmatrix}, \tag{1}$$

where T is a nonsingular $m \times m$ matrix and $\mathbf{T}^0$ is an m-vector. The diagonal elements of T are negative. All other entries of T and the components of $\mathbf{T}^0$ are nonnegative. Moreover

$$T\mathbf{e} + \mathbf{T}^0 = \mathbf{O} \quad \text{where } \mathbf{e} = (1, 1, \ldots, 1)'. \tag{2}$$

The state $m+1$ is absorbing. We require that all other states be transient. The necessary and sufficient condition for this is that the inverse T^{-1} exists. In this case eventual absorption into the state $m+1$ from any initial state $i \in \{1, \ldots, m\}$ is certain.

The vector of initial probabilities is denoted by $(\boldsymbol{\alpha}, \alpha_{m+1})$, where $\boldsymbol{\alpha}$ is an m-vector such that $0 < \boldsymbol{\alpha}\mathbf{e} \leqslant 1$. The probability distribution $F(\cdot)$ of the time till absorption in the Markov chain Q is then easily seen to be

$$F(x) = 1 - \boldsymbol{\alpha} \exp(Tx)\, \mathbf{e}, \quad \text{for } x \geqslant 0. \tag{3}$$

The probability distribution $F(\cdot)$ is said to be of *phase type*. Henceforth this phrase will be rendered as "$F(\cdot)$ is *PH*." The pair $(\boldsymbol{\alpha}, T)$ is called a *representation* of $F(\cdot)$. If $\alpha_{m+1} > 0$, the distribution $F(\cdot)$ has a jump of height α_{m+1} at the origin. All other probability mass is distributed on $(0, \infty)$, according to a density given by

$$\phi(u) = \boldsymbol{\alpha} \exp(Tu)\, \mathbf{T}^0, \quad \text{for } u > 0. \tag{4}$$

The moments μ_k', $k \geqslant 1$, about the origin all exist and are given by the formula

$$\mu_k' = (-1)^k k!\, \boldsymbol{\alpha} T^{-k} \mathbf{e}, \quad \text{for } k \geqslant 1. \tag{5}$$

EXAMPLES: (a) For the exponential distribution with parameter λ, the matrix Q is given by

$$Q = \begin{vmatrix} -\lambda & \lambda \\ 0 & 0 \end{vmatrix}, \qquad \text{and } \alpha_1 = 1,\ \alpha_2 = 0, \tag{6}$$

so that $F(\cdot)$ then has the simple representation $(1, -\lambda)$.

(b) The generalized Erlang distribution obtained by the convolution of m exponential distributions with parameters $\lambda_1, \ldots, \lambda_m$ has as one of its representations the pair $(\boldsymbol{\alpha}, T)$ given by

$$\boldsymbol{\alpha} = (1, 0, \ldots, 0)$$

$$T = \begin{vmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & 0 \\ 0 & -\lambda_2 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & -\lambda_m \end{vmatrix}, \tag{7}$$

with $\mathbf{T}^\circ = (0, \ldots, 0, \lambda_m)'$.

(c) The hyperexponential distribution

$$F(x) = \sum_{i=1}^{m} \alpha_i (1 - e^{-\lambda_i x}), \quad x \geqslant 0, \tag{8}$$

may be represented by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$ and $T = diag(-\lambda_1, -\lambda_2, \ldots, -\lambda_m)$, with $\mathbf{T}^\circ = (\lambda_1, \ldots, \lambda_m)'$.

For any representation $(\boldsymbol{\alpha}, T)$, we can define a matrix Q^* by

$$Q^* = T + T^\circ A^\circ, \tag{9}$$

where T° is an $m \times m$ matrix with m identical column vectors given by the vector $\mathbf{T}^\circ$. The matrix A° is defined to be $(1 - \alpha_{m+1})^{-1}\, diag(\alpha_1, \ldots, \alpha_m)$. The matrix Q^* is easily seen to be a conservative stable matrix. It may be considered to be the transition probability matrix of an m-state Markov chain, which has a close relationship to the probability distribution $F(\cdot)$. As shown in [3], we may always delete possibly superfluous states in the original chain to insure, without loss of generality, that *the matrix Q^* is irreducible.*

The significance of Q^* is as follows. At any time that an absorption occurs in the Markov chain Q, we *instantaneously* perform one or more independent multinomial trials with probabilities $\alpha_1, \ldots, \alpha_{m+1}$, until one of the outcomes $1, \ldots, m$ occurs. The state $i \in \{1, \ldots, m\}$ so obtained is then treated as a new initial state for the Markov chain Q. The process so obtained is an $(m+1)$-state Markov chain in which the state $m+1$ is an instantaneous state. The process obtained by requiring the path functions to be right-hand continuous is again a Markov chain with the m states {1, 2, ..., m} and the infinitesimal generator Q^*. It is routinely verified that the times between the (instantaneous) visits to the state $m+1$ are independent and identically distributed with the common distribution $F(\cdot)$.

The remainder of this paper deals with a unified treatment of renewal processes of phase type. For use in the sequel, we recall the following result, proved in Ref. [3]

THEOREM 1: If the probability distribution $F(\cdot)$ is PH with representation $(\boldsymbol{\alpha}, T)$ then the probability distribution

$$F^*(x) = \frac{1}{\mu_1'} \int_0^x [1 - F(u)]\, du, \quad x \geqslant 0, \tag{10}$$

is PH with the representation $(\boldsymbol{\pi}, T)$ where $\boldsymbol{\pi}$ is the unique probability vector, satisfying

$$\boldsymbol{\pi} Q^* = \boldsymbol{\pi}(T + T^\circ A^\circ) = 0,$$

$$\boldsymbol{\pi} \mathbf{e} = 1. \tag{11}$$

REMARK: The straightforward proof of Theorem 1 is given in Ref. [3]. It is worth pointing out that this result is highly intuitive. In the stationary version of the Markov chain Q^*, π_i is the probability of being in state i, $1 \leqslant i \leqslant m$ at a given time. The PH-distribution with representation $(\boldsymbol{\pi}, T)$ is clearly the probability distribution of the time till the next visit to the instantaneous state $m+1$, i.e., until the next renewal.

COROLLARY 1:

$$(1 - \alpha_{m+1})^{-1} \boldsymbol{\pi} \mathbf{T}^\circ = 1/\mu'_1. \tag{12}$$

PROOF: Formula (11) leads to

$$\boldsymbol{\pi} T = -(1 - \alpha_{m+1}) (\boldsymbol{\pi} \mathbf{T}^\circ) \boldsymbol{\alpha}, \tag{13}$$

and hence

$$\boldsymbol{\pi} \mathbf{e} = -(1 - \alpha_{m+1})^{-1} (\boldsymbol{\pi} \mathbf{T}^\circ) \boldsymbol{\alpha} T^{-1} \mathbf{e} = (1 - \alpha_{m+1})^{-1} (\boldsymbol{\pi} \mathbf{T}^\circ) \mu'_1 = 1. \tag{14}$$

THEOREM 2: The renewal density of an ordinary PH-renewal process is given by

$$\phi(t) = (1 - \alpha_{m+1})^{-2} \boldsymbol{\alpha} \exp[(T + T^\circ A^\circ) t] \mathbf{T}^\circ, \quad \text{for } t \geqslant 0, \tag{15}$$

and we have

$$\lim_{t \to \infty} \phi(t) = \frac{1}{\mu'_1}. \tag{16}$$

PROOF: The quantity $\phi(t)\,dt$ is the expected number of renewals in $[t, t + dt)$. If at least one renewal occurs, the expected number of renewals in $[t, t + dt)$ is given by $(1 - \alpha_{m+1})^{-1}$. The probability that there is at least one renewal in $[t, t + dt)$ is also the probability that, in the Markov chain Q^*, a visit to the instantaneous state $m+1$ occurs during that interval of time. The latter probability is given by

$$(1 - \alpha_{m+1})^{-1} \boldsymbol{\alpha} \exp(Q^* t)\, \mathbf{T}^\circ\, dt, \tag{17}$$

so that (15) follows.

Since Q^* is irreducible, so is the stochastic matrix $\exp(Q^* t)$. In fact, one may prove that $\exp(Q^* t)$ is strictly positive, and therefore aperiodic. The limit matrix of $\exp(Q^* t)$, as $t \to \infty$, is given by Π, with $\Pi_{ij} = \pi_j$, for $1 \leqslant i, j \leqslant m$. It follows that

$$\lim_{t \to \infty} \phi(t) = (1 - \alpha_{m+1})^{-2} \boldsymbol{\alpha} \Pi \mathbf{T}^\circ = (1 - \alpha_{m+1})^{-1} (\boldsymbol{\pi} \mathbf{T}^\circ) = \frac{1}{\mu'_1}, \tag{18}$$

by (12).

Remarks on Computation

We see that $\phi(t) = (1 - \alpha_{m+1})^{-1} \phi_1(t)$, where $\phi_1(\cdot)$ is the renewal density of a PH renewal process with underlying PH-distribution $F_1(\cdot)$ with representation $(\boldsymbol{\beta}, T)$, where $\boldsymbol{\beta} = (1 - \alpha_{m+1})^{-1} \boldsymbol{\alpha}$. We can therefore conveniently assume that $\alpha_{m+1} = 0$, in computing $\phi(t)$.

We only need to solve the system of linear differential equations

$$\mathbf{v}'(t) = \mathbf{v}(t)\ (T + T^{\circ}A^{\circ}), \quad \text{for } t \geqslant 0,$$

$$\mathbf{v}(0) = \boldsymbol{\alpha}, \tag{19}$$

by any one of a large number of numerical methods. For each computed vector $\mathbf{v}(t)$, $\phi(t)$ is then given by

$$\phi(t) = \mathbf{v}(t)\ \mathbf{T}^{\circ}. \tag{20}$$

PH-renewal densities can therefore be readily computed by entirely elementary methods. This is of some practical and pedagogical interest. We also see that the stationary renewal density $\phi_2(t)$ is obtained by choosing the initial state of the chain Q^* according to the probability vector $\boldsymbol{\pi}$. We obtain the expected result

$$\begin{aligned}\phi_2(t) &= (1 - \alpha_{m+1})^{-1}\, \boldsymbol{\pi} \exp(Q^*t)\ \mathbf{T}^{\circ} \\ &= (1 - \alpha_{m+1})^{-1}\, \boldsymbol{\pi}\, \mathbf{T}^{\circ} = 1/\mu_1.\end{aligned} \tag{21}$$

2. THE NUMBER OF RENEWALS IN AN INTERVAL

Let $P_{ij}(n, t)$ be the conditional probability that at time t the Markov chain Q^* is in the state $j \in \{1, \ldots, m\}$ and that n renewals have occurred in the interval $(0, t)$, given that at time $t = 0+$ the chain Q^* was in the state $i \in \{1, \ldots, m\}$. The matrix with entries $P_{ij}(n, t)$ will be denoted by $P(n, t)$. It is then easy to see that the matrices $P(n, t)$ satisfy the recurrence relations

$$P(0, t) = \exp(Tt), \quad \text{for } t \geqslant 0,$$

and

$$\begin{aligned}P(n, t) &= (1 - \alpha_{m+1}) \sum_{\nu=1}^{n} \alpha_{m+1}^{\nu-1} \int_0^t \exp[T(t-u)]\ T^{\circ}A^{\circ}\, P(n-\nu, u)\, du \\ &= (1 - \alpha_{m+1}) \sum_{\nu=1}^{n} \alpha_{m+1}^{\nu-1} \int_0^t P(n-\nu, u)\, T^{\circ}A^{\circ} \exp[T(t-u)]\, du,\end{aligned} \tag{22}$$

for $n \geqslant 1$, $t \geqslant 0$.

The matrix probability generating function

$$P^*(z, t) = \sum_{n=0}^{\infty} P(n, t) z^n, \quad |z| \leqslant 1, \tag{23}$$

is given by the following theorem:

THEOREM 2:

$$P^*(z, t) = \exp\left\{[T + (1 - \alpha_{m+1} z)^{-1} (1 - \alpha_{m+1})\ z\ T^{\circ}A^{\circ}]t\right\}, \quad \text{for } t \geqslant 0. \tag{24}$$

PROOF: The recurrence relations (22) lead to

$$\begin{aligned}P^*(z, t) &= \exp(Tt) + \\ &\quad (1 - \alpha_{m+1}z)^{-1} (1 - \alpha_{m+1}) z \int_0^t \exp[T(t-u)]\ T^{\circ}A^{\circ} P^*(z, u)\, du.\end{aligned} \tag{25}$$

It is well-known that $\exp(-Tt)$ exists and is the inverse of $\exp(Tt)$ [1]. Left-multiplying in (25) by $\exp(-Tt)$ and differentiating the resulting expression with respect to t, we obtain,

$$\exp(-Tt)\,\frac{\partial}{\partial t}\,P^*(z,\,t) = \exp(-Tt)\;T\;P^*(z,\,t)$$

$$+\,(1-\alpha_{m+1}z)^{-1}\,z(1-\alpha_{m+1})\,\exp(-Tt)\;T^{\circ}A^{\circ}P^*(z,\,t), \tag{26}$$

which leads us to the differential equation

$$\frac{\partial}{\partial t}\,P^*(z,\,t) = [T + (1-\alpha_{m+1}z)^{-1}(1-\alpha_{m+1})\,zT^{\circ}A^{\circ}]\;P^*(z,\,t), \tag{27}$$

with the initial condition $P^*(z,\,0) = I$. It now follows directly that $P^*(z,\,t)$ is given by (24).

REMARKS: (a) Formula (24) generalizes the classical formula for the probability generating function of the Poisson distribution. For $m=1$, $T=-\lambda$, $\mathbf{T}^{\circ}=\lambda$, and $\alpha_2=0$, $\alpha_1=1$, we obtain

$$P^*(z,\,t) = \exp(-\lambda + \lambda z)\;t. \tag{28}$$

(b) By a series expansion with respect to z in (27), or by repeating the argument used in proving (26), we see that the matrices $P(n,\,t)$ satisfy the system of linear differential equations.

$$P'(0,\,t) = T\,P(0,\,t) = p(0,\,t)\;T,$$

and

$$P'(n,\,t) = T\,P(n,\,t) + (1-\alpha_{m+1})\sum_{\nu=1}^{n}\alpha_{m+1}^{\nu-1}\;T^{\circ}A^{\circ}P(n-\nu,\,t)$$

$$= P(n,\,t)\;T + (1-\alpha_{m+1})\sum_{\nu=1}^{n}\alpha_{m+1}^{\nu-1}\,P(n-\nu,\,t)\,T^{\circ}A^{\circ}, \tag{29}$$

for $n \geqslant 1$, $t \geqslant 0$, with $p(n,\,0) = \delta_{0n}I$, for $n \geqslant 0$.

Except for very special cases, it will be necessary to solve the system (29) by numerical techniques. We also note that in the case $\alpha_{m+1} = 0$, which occurs in most practical applications, all the preceding expressions simplify considerably.

THEOREM 3: For $t \geqslant 0$, we have

$$\left[\frac{\partial}{\partial t}\,P^*(z,\,t)\,\mathbf{e}\right]_{z=1} = \mu_1'^{-1}\,t\,\mathbf{e} + (1-\alpha_{m+1})^{-1}\,[I-\exp(Q^*t)]\;(\tau^*\Pi - Q^*)^{-1}\,\mathbf{T}^{\circ}, \tag{30}$$

where $\tau^* \geqslant \max[-Q_{ii}]$.

PROOF: Expanding the matrix exponential in Formula (24) and differentiating with respect to z, we obtain

$$\left[\frac{\partial}{\partial t}\,P^*(z,\,t)\right]_{z=1} = (1-\alpha_{m+1})^{-1}\sum_{n=1}^{\infty}\frac{t^n}{n!}\sum_{\nu=0}^{n-1}(Q^*)^{\nu}\,T^{\circ}A^{\circ}(Q^*)^{n-\nu-1}. \tag{31}$$

Right-multiplying on the right by e we obtain, since $Q^*\mathbf{e} = \mathbf{O}$,

$$\left[\frac{\partial}{\partial z} P^*(z, t)\mathbf{e}\right]_{z=1} = (1 - \alpha_{m+1})^{-1} \sum_{n=1}^{\infty} \frac{t^n}{n!} (Q^*)^{n-1} \mathbf{T}^\circ \tag{32}$$

$$= (1 - \alpha_{m+1})^{-1} \int_0^t \exp(Q^*u)\, du\, \mathbf{T}^\circ.$$

The evaluation of the latter integral is of some independent interest, since the (i, j)-entry of the matrix $\int_0^t \exp(Q^*u)\,du$ is the expected amount of time spent in the state j in $(0, t)$, given that the Q^*-chain starts in the state i.

It is easily verified that the matrix

$$P_1 = I + \frac{1}{\tau^*} Q^*, \tag{33}$$

is irreducible, is stochastic, and has the invariant probability vector $\boldsymbol{\pi}$. This implies that the matrix $I - P_1 + \Pi$ is nonsingular [2]. We have the relations

$$I - P_1 + \Pi = \Pi - \frac{1}{\tau^*} Q^*, \qquad \Pi = \Pi\left(\Pi - \frac{1}{\tau^*} Q^*\right). \tag{34}$$

We now see that

$$\int_0^t \exp(Q^*u)\, du \left(\Pi - \frac{1}{\tau^*} Q^*\right) = \sum_{\nu=0}^{\infty} \frac{t^{\nu+1}}{(\nu+1)!} (Q^*)^\nu \left(\Pi - \frac{1}{\tau^*} Q^*\right)$$

$$= \Pi t - \frac{1}{\tau^*} [\exp(Q^*t) - I], \tag{35}$$

from which it follows that

$$\left[\frac{\partial}{\partial z} P^*(z, t)\mathbf{e}\right]_{z=1} = (1 - \alpha_{m+1})^{-1} t\, \Pi\, \mathbf{T}^\circ \tag{36}$$

$$+ (1 - \alpha_{m+1})^{-1} \frac{1}{\tau^*} [I - \exp(Q^*t)] \left(\Pi - \frac{1}{\tau^*} Q^*\right)^{-1} \mathbf{T}^\circ.$$

Using Corollary 1, we obtain (30).

COROLLARY 2: The expected number $H(t)$ of renewals in $(0, t]$ is given by

$$H(t) = (1 - \alpha_{m+1})^{-1} \boldsymbol{\alpha} \left[\frac{\partial}{\partial z} P^*(z, t)\mathbf{e}\right]_{z=1}$$

$$= \mu_1^{-1} t + (1 - \alpha_{m+1})^{-2} \boldsymbol{\alpha} [I - \exp(Q^*t)] (\tau^*\Pi - Q^*)^{-1} \mathbf{T}^\circ, \tag{37}$$

and the expected number of renewals in $[0, t]$ is given by $H(t) + (1 - \alpha_{m+1})^{-1}$.

REMARK: Except for the inverse of the matrix $\tau^*\Pi - Q^*$, which needs to be evaluated only once, the numerical computation of the renewal function again reduces to the computation of $\boldsymbol{\alpha} \exp(Q^*t)$, for various values of t. This may be done by a routine numerical solution of a well-behaved system of linear differential equations with constant coefficients.

3. NUMERICAL METHODS

In a number of applications of PH-distributions, it is necessary to evaluate the matrices $P(n, t)$ for $0 \leqslant n \leqslant N$ and $0 \leqslant t \leqslant T^*$. Usually T^* is given and N needs to be determined so that the components of the vector

$$\sum_{\nu=N+1}^{\infty} P(\nu, T^*)\,\mathbf{e} \tag{38}$$

are all less than a preassigned ϵ.

If N and T^* are known, the system of differential equations

$$P'(0, t) = T\,P(0, t), \tag{39}$$

and

$$P'(n, t) = T\,P(n, t) + \sum_{\nu=1}^{n} \alpha_{m+1}^{\nu-1}\, T^0 A^0\, P(n-\nu, t), \quad \text{for } 1 \leqslant n \leqslant N,$$

with initial conditions $P(0, 0) = I$, $P(n, 0) = 0$, for $1 \leqslant n \leqslant N$, can in most cases of practical interest be solved by a classical procedure such as the Runge-Kutta method. Although the coefficient matrix of the system (39) has a very simple structure, shown in (40), the order in the system is usually sufficiently high that its solution by a "theoretical" method such as the spectral method is impractical and lacking in numerical accuracy.

The coefficient matrix of the system (39) is given by

$$C(N) = \begin{vmatrix} T & \alpha_{m+1} T^0 A^0 & \alpha_{m+1}^2 T^0 A^0 & \cdots & \alpha_{m+1}^{N-1} T^0 A^0 \\ 0 & T & \alpha_{m+1} T^0 A^0 & \cdots & \alpha_{m+1}^{N-2} T^0 A^0 \\ 0 & 0 & T & \cdots & \alpha_{m+1}^{N-3} T^0 A^0 \\ 0 & 0 & 0 & \cdots & \alpha_{m+1}^{N-4} T^0 A^0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & \alpha_{m+1} T^0 A^0 \\ 0 & 0 & 0 & \cdots & T \end{vmatrix} \tag{40}$$

and $C(N)$ is a matrix of order Nm. In view of the special structure of $C(N)$, it is not necessary to store this large matrix in implementing a numerical integration method such as the Runge-Kutta procedure.

The matrix T is also frequently sparse, and this can be used in practice to obtain substantial savings in the number of arithmetic operations involved in solving the system (39).

In most practical cases, the system (39) is also not stiff. Any problems due to stiffness are usually apparent in the entries of the matrix T. If one attempted to approximate, e.g., the degenerate distribution at $c > 0$, by an Erlang distribution of order m and parameter $\lambda = c/m$, one would obtain not only a very high value of m, but also a stiff system of differential equations. Techniques based on distributions of phase type are clearly not suitable in this case. Theoretical results on these problems provide only broad guidelines, and a certain amount of numerical experimentation and the use of build-in accuracy checks appears to be required in practice.

The computer storage requirements depend clearly on the sizes of m and N. In some applications, we are not interested in the matrices $P(n, t)$ themselves, but in some quantity

expressed in terms of them. As discussed in Ref. [7], the applications in queueing theory involve the sequence of matrices A_n, $n \geq 0$, defined by

$$A_n = \int_0^\infty P(n, t)\, d\, K(t), \tag{41}$$

where $K(\cdot)$ is a probability distribution on $[0, \infty)$. The evaluation of an adequate number of terms of the sequence $\{A_n\}$ requires first a truncation of the integral in (41), followed by a numerical integration using the computed matrices $P(n, t)$ at a sufficient number of t-points. We can avoid storing the matrices $P(n, t)$ for a large number of t points by evaluating the integrals in (41) by a "progressive" integration procedure such as Simpson's rule. For special distributions $K(t)$, there may be a substantial gain in computational effort by using an appropriate quadrature method, say Laguerre quadrature, if $K(\cdot)$ is a gamma distribution. These are, however, routine matters of numerical analysis, whose discussion here would only repeat classical material.

We conclude by listing a few practical problems, which can be handled by ad hoc general procedures, but for which a more refined analysis would be welcome.

For a given $T^* > 0$, we need to determine N, so that the conditons stated in (38) holds. More tractably, we can determine N so that the probability of having more than N renewals in $[0, T^*]$ is small. As a crude but useful procedure, we can use known approximations to the mean and the standard deviation $\sigma(T^*)$ of the number of renewals in $[0, T^*]$, and choose, e.g., N to be the smallest integer to exeeced $H(T^*) + 3\sigma(T^*)$. For larger values of T^*, the asymptotic normality of the number of renewals in $[0, T^*]$ can be used for the same purpose. It would be of interest to have a refined analysis of the remainder vector, given in (38), but this does not appear easy.

We also note that the sequence $\{(1 - \alpha_{m+1})^{-1}\, \boldsymbol{\alpha}\, P(k, t)\, \mathbf{e}, k \geq 0\}$ defines a discrete probability density, with parameters $\boldsymbol{\alpha}$, T, and t, which generalizes the Poisson density with parameter λt. There are a variety of particular cases, involving only a small number of parameters, which may be of interest as counting distributions derived from modifications of the Poisson process. If we consider the particular case given in (7), we obtain the probability density

$$p_k = \boldsymbol{\alpha}\, P(k, t)\, \mathbf{e} = \sum_{\nu=0}^{m-1} e^{-\lambda t} \frac{(\lambda t)^{mk+\nu}}{(mk + \nu)!}, \text{ for } k \geq 0. \tag{42}$$

Few particular cases are as tractable analytically as this case. Even for the hyperexponential case, the explicit form of the probabilities p_k is forbiddingly complicated.

REFERENCES

[1] Bellman, R., *Introduction to Matrix Analysis* (McGraw-Hill, New York, 1960).

[2] Kemeny, J., and J. L. Snell, *Finite Markov Chains* (Van Nostrand, Princeton, NJ, 1960).

[3] Neuts, M. F., "Probability Distribution of Phase Type," in *Liber Amicorum Professor Emeritus H. Florin*, pp. 173-206, (Department of Mathematics, University of Louvain, Belgium, 1975).

[4] Neuts, M. F., "Computational Uses of the Method of Phases in the Theory of Queues," Computers and Mathematics with Applications, *1*, 151-166 (1975).

[5] Neuts, M. F.,Computational Problems Related to the Galton-Watson Process," forthcoming in the Proceedings of an Actuarial Research Conference held at Brown University, Providence, RI, in 1975.

[6] Neuts, M. F., "Algorithms for the Waiting Time Distributions under Various Queue Disciplines in the M/G/1 Queue with Service Time Distributions of Phase Type" in *Algorithmic Methods in Probability," TIMS-North Holland Studies in Management Science No. 7* (North Holland, Amsterdam, 1977).

[7] Neuts, M. F., "Markov Chains with Applications in Queuing Theory, Which have a Matrix-Geometric Invariant Probability Vector," Advances in Applied Probability, *10*, 185-212, (1978).

TECHNIQUES FOR ESTABLISHING ERGODIC AND RECURRENCE PROPERTIES OF CONTINUOUS-VALUED MARKOV CHAINS

G. M. Laslett

Division of Mathematics and Statistics
Commonwealth Scientific and Industrial Organization
Melbourne, Australia

D. B. Pollard

Statistics Department, IAS
Australian National University
*Canberra, Australia**

R. L. Tweedie

Division of Mathematics and Statistics
Commonwealth Scientific and Industrial Organization
Canberra, Australia

ABSTRACT

We present techniques for classifying Markov chains with a continuous state space as either ergodic or recurrent. These methods are analogous to those of Foster for countable space chains. The theory is presented in the first half of the paper, while the second half consists of examples illustrating these techniques. The technique for proving ergodicity involves, in practice, three steps: showing that the chain is irreducible in a suitable sense; verifying that the mean hitting times on certain (usually bounded) sets are bounded, by using a "mean drift" criterion analogous to that of Foster; and finally, checking that the chain is such that bounded mean hitting times for these sets does actually imply ergodicity.

The examples comprise a number of known and new results: using our techniques we investigate random walks, queues with waiting-time-dependent service times, dams with general and random-release rules, the s-S inventory model, and feedback models.

1. INTRODUCTION

Since the introduction of the embedded chain method by Kendall [9], Markov chain analysis has been recognized as an important tool in many branches of operations research. The most closely studied chains have undoubtedly been those which are integer-valued. This is so partly because a very complete theory exists for analyzing such chains, but a contributing factor has also been the formulation (since the results of Foster [6]) of readily verifiable criteria for classifying integer-valued chains as ergodic or as recurrent. Recent developments in Markov chain theory now enable these criteria to be extended to cover the classification of

*Currently at Department of Statistics, Yale University, New Haven, Connecticut.

chains taking values in more general spaces. Tweedie [19] has given a review of this general theory.

In this paper the special case of Markov chains whose state space is the real line, or a closed subset of the real line, will be considered. We shall develop techniques for proving the existence of unique stationary distributions, and for showing that the n-step transition probabilities converge to this stationary distribution. These techniques enable us to classify a number of the continuous-valued chains which occur naturally in operations research; applications to random walks, waiting times for queues, capacities of dams, inventory problems, and feedback chains are given.

2. CONTINUOUS-VALUED MARKOV CHAINS

In order to motivate the description of continuous-valued chains, we begin with a review of the corresponding more familiar concepts for the integer-valued case.

The evolution of a (time homogeneous) integer-valued chain is described by its transition matrix $P = [P(i, j)]$ defined by

$$P(i, j) = \Pr\{X_n = j | X_{n-1} = i\}.$$

Suppose that the chain is *irreducible*, i.e., for every pair (i, j) there exists an n such that

$$Pr\{X_n = j | X_0 = i\} > 0.$$

In the analysis of such chains, a question of fundamental interest concerns the existence of a unique stationary distribution $\{\pi(k)\}$: that is, a distribution satisfying

$$\pi(k) = \sum_j \pi(j)\, P(j, k) \tag{2.1}$$

for all k. If such a distribution exists $\{X_n\}$ is said to be *ergodic* (or positive recurrent).

A concept weaker than ergodicity is *recurrence*. Consider the hitting times

$$T_k = \inf\{n > 0 : X_n = k\}.$$

The chain $\{X_n\}$ is said to be recurrent if these variables are proper for any starting point X_0; that is, if

$$Pr\{T_k < \infty | X_0 = j\} = 1$$

for all j and k. It is well known (Feller, Ref. [7], Chapter XV) that for an irreducible chain recurrence is equivalent to the single condition

$$Pr\{T_0 < \infty | X_0 = 0\} = 1.$$

Further, the chain is ergodic if and only if

$$E[T_0 | X_0 = 0] < \infty \tag{2.2}$$

and the stationary distribution is then given by

$$\pi(k) = (E[T_k | X_0 = k])^{-1} > 0$$

for each k. In all but the simplest cases it is impossible to solve (2.1) directly, or to find the distribution of T_0 conditional on $X_0 = 0$ explicitly enough to check for recurrence or ergodicity. However Foster [6], and later Mauldon [12] and Pakes [14], have given sufficient conditions for recurrence and ergodicity of irreducible integer-valued chains. We extend these conditions to continuous-valued chains.

Consider a time-homogeneous Markov chain $\{X_n\}$, with state space $\mathfrak{X}$, which we will usually assume to be a closed (but not necessarily bounded) subset of the real line $(-\infty, \infty)$. (All our results can be given virtually unaltered when $\mathfrak{X}$ is taken as a closed subset of a higher dimensional Euclidean space, but for ease of exposition we generally restrict ourselves to the one dimensional case here.) The evolution of the chain is described in one dimension by the collection of distribution functions

$$F_x(y) = Pr\{X_{n+1} \leqslant y \mid X_n = x\};$$

however, we will find it easier to work with the corresponding measures

$$P(x, A) = Pr\{X_{n+1} \in A \mid X_n = x\}$$

induced in one dimension by the distributions F_x; that is, if $A = (a, b]$, then $P(x, A) = F_x(b) - F_x(a)$, and $P(x, \cdot)$ is then extended to all the Borel subsets of the real line in the usual way. We assume, in order that the chain be well-defined, that for each $A \in \mathcal{F}$ (the σ-field of Borel subsets of $\mathfrak{X}$) the function $P(\cdot, A)$ is measurable, and for each x, $P(x, \cdot)$ is a probability measure on $\mathcal{F}$.

As in the discrete case, to classify such a chain we need a notion of irreducibility.

DEFINITION: We say $\{X_n\}$ is *ϕ-irreducible* if there exists a nonzero measure ϕ on F such that, for any $x \in \mathfrak{X}$ and $A \in \mathcal{F}$ with $\phi(A) > 0$, there is an n for which $P^n(x, A) > 0$.

When $\mathfrak{X}$ is countable, taking ϕ as counting measure leads us back to the usual concept of irreducibility. For general $\mathfrak{X}$, ϕ need not have any atoms (ϕ will often be, for example, Lebesgue measure) but if it does have atoms then the analysis is greatly simplified (see Section 10).

We shall call a ϕ-irreducible chain $\{X_n\}$ *ergodic* if it has a unique stationary distribution, i.e. a probability measure π on $\mathcal{F}$ satisfying, for every $A \in \mathcal{F}$,

$$\pi(A) = \int_{\mathfrak{X}} \pi(dx)\, P(x, A). \tag{2.3}$$

If $\{X_n\}$ is ϕ-irreducible then (Tweedie, Ref. [19], Section 4) there is *at most* one stationary distribution; if further $\{X_n\}$ is ergodic then the n-step transition probabilities converge to the stationary distribution π in the strong Cesaro sense that

$$\sup_{A \in F} \left| \frac{1}{n} \sum_{m=1}^{n} P^m(x, A) - \pi(A) \right| \to 0 \tag{2.4}$$

for π-almost all x.

Thus ergodicity has connotations for continuous state spaces similar to those for the discrete case. Moreover, there is again a close relationship between ergodicity and the finiteness of the means of the hitting times

$$T_A = \inf\{n > 0: X_n \in A\}.$$

This connection is not as simple as the necessary and sufficient condition given by (2.2) for discrete chains. But it can be shown that, given certain conditions on the chain, ergodicity is a consequence of:

$$\sup_{x \in A} E[T_A \mid X_0 = x] < \infty, \tag{2.5}$$

provided A is one of a certain class of sets determined by the preliminary conditions satisfied by the chain. One of the main aims of this paper is to detail the conditions which lead to a useful class of sets in this context. For example, we shall show that ϕ-irreducibility plus certain con-

tinuity constraints on the transition probabilities implies that we need only verify (2.5) for a single bounded set A of positive ϕ-measure, in order to prove that the chain is ergodic. In general, we call any set A such that (2.5) is a sufficient condition for ergodicity of $\{X_n\}$ a *test set* (*for ergodicity*) for that chain.

In the next four sections we describe some suitable test sets for four important classes of Markov chains which occur in operations-research models. For such chains then, ergodicity can be proved by finding a test set satisfying (2.5). This will always be achieved by applying the following result.

THEOREM 2.1 (Tweedie, Ref. [19]), Theorem 4.1): Let g be a nonnegative measurable function on $\mathfrak{X}$. If for some $\epsilon > 0$ and $A \in \mathcal{F}$,

$$\int_{\mathfrak{X}} P(x, dy)\, g(y) \leqslant g(x) - \epsilon \quad \text{for } x \in A^c, \tag{2.6}$$

$$\text{i.e., } E[g(X_1)|X_0 = x] \leqslant g(x) - \epsilon \quad \text{for } x \in A^c,$$

then we have the bounds

$$E[T_A|X_0 = x] \leqslant g(x)/\epsilon \quad \text{for } x \in A^c, \tag{2.7}$$

and

$$E[T_A|X_0 = x] \leqslant 1 + \int_{A^c} P(x, dy)\, g(y)/\epsilon \quad \text{for } x \in A. \tag{2.8}$$

A very common choice for the test function g is $g(x) = x$ when the space $\mathfrak{X} = [0, \infty)$. This will be the case for the examples of Sections 10, 11, 12, and 14. Thus it seems worthwhile to show explicitly how Theorem 2.1 can be used for that choice of g.

THEOREM 2.2: Suppose $\mathfrak{X} = [0, \infty)$ and that there exist $\epsilon > 0$, $M < \infty$ and a bounded $A \in \mathcal{F}$ such that

$$E[X_1|X_0 = x] \leqslant x - \epsilon \quad \text{for } x \in A^c \tag{2.9}$$

and

$$E[X_1|X_0 = x] \leqslant M \quad \text{for } x \in A. \tag{2.10}$$

Then

$$\sup_{x \in A} E[T_A|X_0 = x] < \infty.$$

PROOF: Apply Theorem 2.1 with $g(x) = x$, noticing that x/ϵ is bounded on A and that the right-hand side of (2.8) is bounded by $1 + M/\epsilon$.

In most of the cases which we shall consider the space $\mathfrak{X}$ is $[0, \infty)$ and the test set will be a bounded interval $[0, \beta]$. We can then interpret (2.9) as "mean drift towards" this test set. With this sort of interpretation, many of our results are more intuitively meaningful: they can be seen as providing some delineation of the class of chains for which mean drift towards "reasonable" sets does in fact imply ergodicity.

The discussion above, and the method we advocate, can be summarized in the following:

ERGODICITY TECHNIQUE: To prove ergodicity for a given Markov chain $\{X_n\}$, carry out the following three steps:

STEP I. Identify a suitable ϕ and show $\{X_n\}$ is ϕ-irreducible for this ϕ.

STEP II. Identify possible test sets for the chain.

STEP III. Apply Theorem 2.1 (or Theorem 2.2) to one of these test sets to prove boundedness of the mean hitting times as specified by (2.5).

The reader who is primarily interested in applications could now turn straight to Section 9, which is the beginning of the examples segment of this paper. Of course some back-referencing would then be required: to Sections 3-6 to see how the test sets have been identified; to Sections 7 for ϕ-irreducibility; and possibly to Section 8 as well, where we describe a concept of recurrence which generalizes the one for discrete chains.

3. TEST SETS WHEN ϕ HAS AN ATOM

Sometimes ϕ can be chosen to have an atom at some point α, i.e. $\{\alpha\}$ can be reached from every point in the state space (see Sections 10 and 11).

THEOREM 3.1 (Tweedie, Ref. [19], Section 5): If ϕ has an atom at α then $\{\alpha\}$ is a test set for any ϕ-irreducible chain.

From this simple result we obtain the more usable condition:

THEOREM 3.2 (Tweedie, Ref. [19], Section 5): If ϕ has an atom at α then a set B containing α is a test set if for some integer N and some $\delta > 0$

$$\max_{n \leqslant N} P^n(y, \{\alpha\}) \geqslant \delta$$

for every $y \in B$.

If $\mathcal{X} = \{0, 1, 2, \ldots,\}$ and $\{X_n\}$ is irreducible, then Theorem 3.1 implies that $\{0\}$ is a test set; Theorem 3.2 covers the result that $\{0, 1, \ldots N\}$ is a test set for any finite N. See Section 10.

4. TEST SETS WHEN THE CHAIN IS WEAKLY CONTINUOUS

In the absence of the discrete type of behavior of Section 3 we can resort to a continuity condition on the chain to identify some useful test sets.

DEFINITION: $\{X_n\}$ is said to be *weakly continuous* if for every bounded continuous real function f on $\mathcal{X}$

$$Pf(x) = \int_{\mathcal{X}} P(x, dy)\, f(y)$$

is also bounded and continuous. Equivalently: if $x_n \to x$ then $P(x_n, \cdot) \to P(x, \cdot)$ in distribution, i.e., $F_{x_n}(y) \to F_x(y)$ at every continuity point of the latter in one dimension.

THEOREM 4.1 (Tweedie, Ref. [19], Section 5): If $\{X_n\}$ is a weakly continuous ϕ-irreducible chain then any bounded set of positive ϕ-measure is a test set for that chain.

Proving weak continuity in order to use Theorem 4.1 may often be difficult if the chain has a complicated structure. Since we feel the weak continuity condition will prove the most frequently used in practice, we give in the remainder of this section a routine for proving weak continuity which reduces the evaluation of quite complicated models to a series of relatively simple steps.

In many models the chain under consideration may be regarded as a secondary process, derived from a primary process about which we are prepared to make assumptions. In particular, we aim to carry weak continuity of the primary process over to the secondary one. That is, suppose the transition of the chain $\{X_n\}$ (with transition probabilities $P(x, \cdot)$) from x to x' can be decomposed into a sequence of simpler transitions: from x to y_1 according to a transition law $Q_1(x, \cdot)$; then from y_1 to y_2 according to law $Q_2(y_1, \cdot)$; ...; and finally from y_{k-1} to x' according to the law $Q_k(y_{k-1}, \cdot)$. Then weak continuity of each of the $Q_i(y_{i-1}, \cdot)$'s results in the same property for the $P(x, \cdot)$. See Section 11 for an example of such a decomposition. To utilize this concept we give the following simple results.

THEOREM 4.2: If the transition law $P(x, \cdot)$ can be decomposed as

$$P(x, A) = \int \dots \int Q_1(x, dy_1)\, Q_2(y_1, dy_2) \dots Q_k(y_{k-1}, A) \tag{4.1}$$

where each $Q_i(y_{i-1}, \cdot)$ is weakly continuous, then $P(x, \cdot)$ is also weakly continuous.

PROOF: If f is bounded and continuous, and $Q_k(y_{k-1}, \cdot)$ is weakly continuous, then $Q_k f(y_{k-1})$ is continuous; by induction, from (4.1), $Pf(x) = Q_1 Q_2 \dots Q_k f(x)$ is continuous, which is the desired result.

When applying this theorem we usually make the transition probabilities $Q_j(y_{j-1}, \cdot)$ as simple as possible (see Section 11). Here are some of the simple forms of transition probabilities for which weak continuity can be proved fairly easily:

(i) A transition law $R(x, \cdot)$ on $\mathfrak{X}$ will be called *degenerate* (at h) if there is a function h such that $R(x, \cdot)$ is concentrated at $h(x)$ for every $x \in \mathfrak{X}$. For example, in Section 11 we shall use $h(x, y, z) = [(x + y - z)^+, y, z]$.

(ii) For a vector process, the transition law $R(x, \cdot)$ may affect only a subset of the coordinates of x. Formally, suppose $\mathfrak{X}$ is a product $\mathfrak{X}_1 \times \mathfrak{X}_2$ of two spaces of lower dimension, and suppose also that $R(x, \cdot)$ leaves the $\mathfrak{X}_2$ coordinate unchanged in the sense that (writing $x = [x_1, x_2]$ in the obvious manner) there is a probability measure $R^*([x_1, x_2], \cdot)$ on $\mathfrak{X}_1$ satisfying

$$\int_{\mathfrak{X}} R([x_1, x_2], d[y_1, y_2])\, f(y_1, y_2) = \int_{\mathfrak{X}_1} R^*([x_1, x_2], dy_1)\, f(y_1, x_2) \tag{4.2}$$

for every bounded measurable f on $\mathfrak{X}$. In this case we say that $R(x, \cdot)$ *acts only on the* $\mathfrak{X}_1$ *coordinate of* $\mathfrak{X}$. For example, the transitions in Theorem 11.1 are of this form.

(iii) When $\mathfrak{X} = \mathfrak{X}_1 \times \mathfrak{X}_2$, one of the coordinate processes may form a Markov chain in its own right. That is, if $X_n = [X_{n1}, X_{n2}]$, then $\{X_{n1}\}$ (say) may be a Markov chain. In this case the transition law of the X_{n1} chain is given by

$$R_1(x_1, A) = R([x_1, x_2], A \times \mathfrak{X}_2) \tag{4.3}$$

for any $x_2 \in \mathfrak{X}_2$. We shall identify this situation by saying that *the projection on the* $\mathfrak{X}_1$ *coordinate space is Markovian.* Such is the case with the waiting times of Section 11.

As might be expected, weak continuity for transitions of any of these forms is easier to establish.

THEOREM 4.3: Suppose $R(x, \cdot)$ is a transition law on $\mathfrak{X}$. Then

(i) if $R(x, \cdot)$ is degenerate at h, then it is weakly continuous if and only if h is continu-

(ii) if $\mathfrak{X} = \mathfrak{X}_1 \times \mathfrak{X}_2$ and $R(x, \cdot)$ acts only on the $\mathfrak{X}_1$ coordinate of $\mathfrak{X}$, then it is weakly continuous if and only if

$$R^*g([x_1, x_2]) = \int_{\mathfrak{X}_1} R^*([x_1, x_2], dy_1)\, g(y_1)$$

is continuous for each bounded continuous g on $\mathfrak{X}_1$;

(iii) if $\mathfrak{X} = \mathfrak{X}_1 \times \mathfrak{X}_2$ and the projection onto the $\mathfrak{X}_1$ coordinate space is Markovian (as in (4.3)), then the weak continuity of $R([x_1, x_2], \cdot)$ implies the weak continuity of $R_1(x_1, \cdot)$;

(iv) $R(x, \cdot)$ is weakly continuous if and only if $Rf(x)$ is a continuous function of x for every bounded *uniformly* continuous f on $\mathfrak{X}$.

PROOF: (iv) This follows directly from Theorem 2.1 (ii) of Billingsley [2].

(i) $Rf(x) = f[h(x)]$ which is continuous in x for every bounded continuous f if and only if h is continuous.

(ii) For any bounded continuous g on $\mathfrak{X}_1$, the function $f(x_1, x_2) = g(x_1)$ is bounded and continuous on $\mathfrak{X}$. Since $Rf(x_1, x_2) = R^*g(x_1, x_2)$, the continuity of R^*g follows from that of Rf.

Conversely, suppose R^*g is continuous for every bounded continuous g on $\mathfrak{X}_1$. Let f be bounded and uniformly continuous on $\mathfrak{X}$. If $x_n = [x_{n1}, x_{n2}] \to x_0 = [x_{01}, x_{02}]$ as $n \to \infty$ then

$$|Rf(x_n) - Rf(x_0)| \leqslant \int R^*(x_n, dy_1)|\, f(y_1, x_{n2}) - f(y_1, x_{02})|$$
$$+ |\int R^*(x_n, dy_1)\, f(y_1, x_{02}) - \int R^*(x_0, dy_1)\, f(y_1, x_{02})|.$$

The first term tends to zero as $x_{n2} \to x_{02}$ since f is uniformly continuous; the second term tends to zero because, for fixed x_{02}, $g(y_1) = f(y_1, x_{02})$ is continuous and R^*g is then continuous by assumption.

(iii) If g is bounded and continuous on $\mathfrak{X}_1$ then $f(x_1, x_2) = g(x_1)$ is bounded and continuous on $\mathfrak{X} = \mathfrak{X}_1 \times \mathfrak{X}_2$. Hence

$$\int R_1(x_1, dy_1)\, g(y_1) = \int R([x_1, x_2], d[y_1, y_2])\, f(y_1, y_2)$$

is a bounded continuous function (not depending on x_2).

5. TEST SETS FOR CHAINS WITH WEAKLY CONTINUOUS COMPONENTS

Suppose $\{\tilde{P}(x, \cdot)\}$ is a collection of *substochastic* transition measures, i.e. for each x, $\tilde{P}(x, \cdot)$ is a measure on $\mathcal{F}$ with $\tilde{P}(x, \mathfrak{X}) \leqslant 1$, and $\tilde{P}(\cdot, A)$ is a measurable function for each $A \in \mathcal{F}$. The definition of the iterates $\tilde{P}^n(x, A)$, and the concepts of ϕ-irreducibility and weak continuity of the $\{\tilde{P}(x, \cdot)\}$, are entirely analogous to those for the case of stochastic transition laws. In the same way, all the methods for proving weak continuity given in Section 4 carry over to the substochastic case.

Now if $\{P(x, \cdot)\}$ is the family of transition probabilities for a Markov chain $\{X_n\}$, then $\{\tilde{P}(x, \cdot)\}$ is said to be a *component* of $\{P(x, \cdot)\}$ if

$$P(x, A) \geqslant \tilde{P}(x, A) \quad \text{for every } x \in \mathfrak{X},\ A \in \mathcal{F}.$$

The existence of suitably well-behaved components can be used to identify test sets.

THEOREM 5.1 (Tweedie, Ref. [19], Section 5): If the chain $\{X_n\}$ has a weakly continuous, ϕ-irreducible component, then any bounded set of positive ϕ-measure is a test set for the chain.

The use of components is illustrated in Section 12.

Finally, we note a simple sufficient condition for $\{X_n\}$ to have a weakly continuous ϕ-irreducible component. Suppose

$$P(x, \cdot) = \alpha(x)\ P_1(x, \cdot) + [1 - \alpha(x)]\ P_2(x, \cdot),$$

where $\alpha(x)$ is a continuous function with $0 < \alpha(x) \leqslant 1$ for all x. If $P_1(x, \cdot)$ is ϕ-irreducible and weakly continuous, then $\tilde{P}(x, \cdot) = \alpha(x)\ P_1(x, \cdot)$ is a weakly continuous ϕ-irreducible component of $\{X_n\}$, no matter how badly behaved $P_2(x, \cdot)$ is.

6. TEST SETS DEFINED BY SUBINVARIANT MEASURES

Our final method for finding test sets is the hardest to use in practice, but it will prove necessary in Section 13 since the results of Sections 3, 4, and 5 are inadequate for handling the inventory model.

If $\{X_n\}$ is a ϕ-irreducible chain, then it can be shown (Tweedie, Ref. [18]) that there *exists at least one nontrivial σ-finite measure μ, with $\mu >> \phi$, such that*

$$\mu(A) \geqslant \int_{\mathcal{X}} \mu(dy)\ P(y, A) \quad \text{for all } A \in \mathcal{F}. \tag{6.1}$$

Such a μ is called a subinvariant measure.

Unfortunately, to go beyond the mere existence of μ is often tantamount to finding an invariant measure (i.e. one for which the inequality in (6.1) can be replaced by an equality), which is close to proving ergodicity directly. Nevertheless, the following is useful on occasion.

THEOREM 6.1 (Tweedie, Ref. [19], Chapter 5): Let $\{X_n\}$ be a ϕ-irreducible chain with some subinvariant measure μ. Then any $A \in \mathcal{F}$ with $0 < \mu(A) < \infty$ is a test set.

7. PROVING ϕ-IRREDUCIBILITY

Without ϕ-irreducibility the chain may not be classifiable at all, so it seems worthwhile to give some guidelines for establishing irreducibility. As with integer-valued chains though, it may often be necessary merely to assume ϕ-irreducibility, or to give conditions which are "grossly sufficient" for ϕ-irreducibility (see Section 11). There are two cases where irreducibility can be easily checked:

(i) If there is some point $\alpha \in \mathcal{X}$ which can be reached with positive probability from every point in the space, then we can take ϕ to consist of a single atom at α. See Section 11 for example.

(ii) If $\{X_n\}$ has a ϕ-irreducible component (see Section 5), then it is itself also ϕ-irreducible. Although trivial, this observation is sometimes quite useful since it may be easier to work with the iterates of some component of the chain, rather than with the $P^n(x, \cdot)$'s themselves.

8. RECURRENCE FOR CONTINUOUS-VALUED CHAINS

In the continuous case there are various possible definitions of recurrence: see Tweedie, Ref. [19], Section 3. Here we shall call a ϕ-irreducible chain $\{X_n\}$ *recurrent* if there is a ϕ-null set N such that for all $x \notin N$,

$$Pr\{T_B < \infty | X_0 = x\} = 1 \tag{8.1}$$

for *every* $B \in F$ with $\phi(B) > 0$. The null set N occurs in our definition because one can prove that, if $\{X_n\}$ is not recurrent in this way, then it is transient in a natural manner (Tweedie, Ref. [19], Section 3); such a classification fails if we demand that (8.1) hold for *all* x and all B with $\phi(B) > 0$.

Because a dichotomy between transience and recurrence exists, we can again hope to find individual sets that will help classify the chain itself. If the validity of (8.1) for a particular B implies the recurrence of the chain $\{X_n\}$, then we say that B is a *recurrence test set* for that chain. The next two results then give, respectively, criteria for identifying recurrence test sets, and a criterion for investigating recurrence using these sets.

THEOREM 8.1 (Tweedie, Ref. [19], Section 5): Each of the test sets (for ergodicity) identified in Theorems 3.1, 3.2, 4.1, 5.1 and 6.1 is also a recurrence test set for the same class of chains.

THEOREM 8.2 (Tweedie, Ref. [19], Section 10): Suppose $\{X_n\}$ is ϕ-irreducible. Then a sufficient condition for recurrence of $\{X_n\}$ is the existence of a non-negative measurable function g and an increasing sequence $\{A_n\}$ of recurrence test sets for $\{X_n\}$ satisfying

$$\int_{\mathcal{X}} P(x, dy)\, g(y) \leqslant g(x) \quad \text{for } x \in A_0^c \tag{8.2}$$

$$\{y : g(y) \leqslant n\} \subseteq A_n, \quad n = 1, 2, \ldots \tag{8.3}$$

Whilst (8.2) is similar to the condition (2.6) for bounded mean return times, (8.3) is a restriction on the type of "test function" that can be used in proving recurrence. In general, however, (8.3) is not overrestrictive. We conclude by giving the most common form of use of Theorems 8.1 and 8.2, based on Theorem 5.1 and the test function $g(x) = x$.

THEOREM 8.3 Suppose $\mathcal{X} = [0, \infty)$ and $\{X_n\}$ is ϕ-irreducible with a weakly continuous component. Then a sufficient condition for $\{X_n\}$ to be recurrent is the existence of a constant β such that

$$\int_{\mathcal{X}} P(x, dy)\, y \leqslant x \quad \text{for } x \geqslant \beta. \tag{8.4}$$

In our examples we concentrate on proving ergodicity rather than recurrence. There is a pragmatic reason for this: the ergodicity criteria of Theorems 2.1 and 2.2 are often, rather surprisingly, more easily verified than the recurrence criteria of Theorems 8.1 and 8.2. For example, we often impose a condition which implies that, as $x \to \infty$,

$$\limsup \left[\int_{\mathcal{X}} P(x, dy)\, y - x \right] < 0,$$

which immediately gives (2.9) for large enough x; the analogous condition to give recurrence would seem to be

$$\limsup \left[\int_{\mathcal{X}} P(x, dy)\, y - x \right] \leqslant 0,$$

but this does not give (8.4). Hence we counsel some care in writing down the "natural" extension of ergodicity conditions as recurrence conditons. However, as it is usually quite simple to adapt the conditions for ergodicity to cover the recurrence case, we shall generally leave the formulation of these extensions to the reader.

9. INTRODUCTION TO THE EXAMPLES

The three steps which constitute the basis of our technique for proving ergodicity were listed near the end of Section 2. We now illustrate the ways in which this technique can be applied, by proving a number of ergodicity results for a variety of Markov chain situations. We do not pretend that these results are in any sense optimal; indeed, most of the conclusions could be further refined, if methods suited to the particular example were used (see Kiefer and Wolfowitz [10], or Loynes [11]). Rather, we wish to demonstrate the essential simplicity of our three-step method. In each case the proof is arranged so as to emphasise this point.

10. RANDOM WALKS

One of the simplest examples of a continuous-valued Markov chain is the random walk $\{X_n\}$ on $[0, \infty)$ defined by

$$X_n = (X_{n-1} + Y_n)^+ \tag{10.1}$$

where $\{Y_n\}$ is a sequence of independent, identically distributed $(-\infty, \infty)$-valued random variables. The next result is extremely well-known; we give it because it is a straightforward example of the techniques espoused above.

THEOREM 10.1: If $E(Y_k) < 0$ then $\{X_n\}$ is ergodic.

PROOF: *I. ϕ-irreducibility:* We show that ϵ_0, the point mass at 0, is a suitable irreducibility measure. Choose a $\delta > 0$ such that $\Pr\{Y_k \leqslant -\delta\} = \gamma > 0$. Then for $n \geqslant x\delta^{-1}$,

$$Pr\{X_n = 0 | X_0 = x\} \geqslant \gamma^n > 0.$$

II. Test sets: Any set of the form $[0, \beta]$ is a test set. This can be proved either by showing that the chain is weakly continuous or by applying Theorem 3.2; if $N \geqslant \beta\delta^{-1}$ then, as above,

$$\max_{n \leqslant N} P^n(y, \{0\}) \geqslant \gamma^N \text{ for every } y \in [0, \beta].$$

III. Boundedness of hitting times: Apply Theorem 2.2.

$$\begin{aligned} E[X_1 - X_0 | X_0 = x] &= E[\max\{Y_1, -x\}] \\ &\to E(Y_1) \text{ as } x \to \infty, \text{ by dominated convergence} \\ &< 0, \end{aligned}$$

thus $E(X_1 | X_0 = x) \leqslant x - \epsilon$ for large enough x.

11. QUEUES

Some of the standard queueing results follow from the preceding results for random walks. In this section we consider a slightly more complicated situation: a queue with waiting-time-dependent service times. Such a queue was studied by Callahan [3], but in a rather

artificial discrete setting. Tweedie [20] discussed the continuous state space version of this model; our technique for proving ergodicity can be used to weaken his conditions. Tweedie [21] also discusses this model in more detail using a stronger version of our Theorem 2.1.

Let the interarrival times $\{T_n\}$ of customers in a single-server queue be independent and identically distributed with mean $\mu < \infty$. Let W_n denote the waiting time of the n^{th} customer, and S_n the service time. The random variables $\{S_n\}$ are assumed to be conditionally independent of each other and of the T_n's, given the relevant waiting times, but the distribution of S_n does depend on the waiting time W_n:

$$Pr\{S_n \in A \mid W_n = \omega\} = \gamma_\omega(A).$$

The waiting times $\{W_n\}$ are defined recursively by

$$W_n = (W_{n-1} + S_{n-1} - T_n)^+; \tag{11.1}$$

$\{W_n\}$ is thus a Markov chain with state space $[0, \infty)$. Proving the ϕ-irreducibility of $\{W_n\}$ is difficult; we give here two "grossly sufficient" sets of conditions which apply to two different types of ϕ. Both of these conditions are weaker than those proposed by Callahan [3] for discrete state space chains, or Tweedie [20] for the continuous state space.

(a) For all $\omega \geqslant 0$, and some $\delta > 0$ and $\gamma > 0$, $Pr\{S_{n-1} - T_n < -\delta \mid W_{n-1} = \omega\} \geqslant \gamma$. This makes $\{W_n\}$ ϵ_o-irreducible.

(b) There exist $N > 0$, $\eta > 0$ such that $Pr\{S_{n-1} - T_n > \eta \mid W_{n-1} = \omega\} > 0$ for $\omega \leqslant N$; and the distribution of $S_{n-1} - T_n$ conditional on $W_n = \omega$ has a positive density on $(-\eta, 0)$ if $\omega > N$. This makes $\{W_n\}$ ϕ-irreducible with ϕ as Lebesgue measure on $(N - \eta, N + \eta)$.

THEOREM 11.1: Suppose $\{W_n\}$ is ϕ-irreducible and that (i) $\gamma_\omega(\cdot)$ is weakly continuous in ω;

$$\text{(ii) } \sup_\omega E[S_n \mid W_n = \omega] < \infty \tag{11.2}$$

and

$$\limsup_{\omega \to \infty} E[S_n \mid W_n = \omega] < \mu. \tag{11.3}$$

Then $\{W_n\}$ is ergodic.

PROOF: I. *ϕ-irreducibility:* has already been assumed.

II. Test sets: We prove that $\{W_n\}$ is weakly continuous, so that $[0, \beta]$ is a test set for all large enough values of β (Theorem 4.1).

Decompose the trivariate chain $X_n = (W_n, S_{n-1}, T_n)$ as

$$X_n = (W_n, S_{n-1}, T_n) \to (W_n, S_{n-1}, T_{n+1}) \tag{11.4}$$

$$\to (W_n, S_n, T_{n+1}) \tag{11.5}$$

$$\to [(W_n + S_n - T_{n+1})^+, S_n, T_{n+1}] \tag{11.6}$$

$$= X_{n+1}.$$

Then (11.6) is degenerate at the continuous function $h(x, y, z) = [(x + y - z)^+, y, z]$, and so is weakly continuous by Theorem 4.3 (i); (11.4) is trivially weakly continuous since T_{n+1} is independent of T_n (and X_n) — use Theorem 4.3(ii); and (11.5) is weakly continuous by

assumption — use hypothesis (i) and Theorem 4.3(ii). Thus the chain $\{X_n\}$ is weakly continuous, by Theorem 4.2, and hence the projection $\{W_n\}$ is also weakly continuous, by Theorem 4.3(iii).

III. Boundedness of hitting times: Apply Theorem 2.2 with $A = [0, \beta]$ for a large enough value of β.

$$E[W_1 - W_0 | W_0 = \omega] = E[(\omega + S_0 - T_1)^+ - \omega | W_0 = \omega]$$
$$= E[\max\{S_0 - T_1, -\omega\} | W_0 = \omega].$$

This last term is certainly greater than

$$E[S_0 - T_1 | W_0 = \omega],$$

but exceeds it only by an amount

$$E[\max\{0, T_1 - S_0 - \omega\} | W_0 = \omega] \leqslant E[(T_1 - \omega)^+]$$
$$\to 0 \text{ as } \omega \to \infty.$$

Thus it suffices to show that

$$\limsup_{\omega \to \infty} E[S_0 - T_1 | W_0 = \omega] < 0.$$

This is equivalent to (11.3).

The proof of ergodicity for the usual waiting times in the GI/G/1 queue is contained in Theorem 11.1. For this special case one could also use Theorem 3.2 for establishing that bounded sets are test sets for ergodicity.

The assumption that $\gamma_\omega(\cdot)$ is weakly continuous is a natural extension of the case where S_n is independent of W_n; however it does not cover the case of deterministic service times, i.e., where $\gamma_\omega(\cdot)$ is concentrated at the point $\gamma(\omega)$, unless $\gamma(\omega)$ is a continuous function. We can handle some discontinuous $\gamma(\omega)$'s by using components.

THEOREM 11.2: Suppose $\gamma_\omega(\cdot)$ is concentrated at $\gamma(\omega)$, and that $\gamma(\omega)$, although not necessarily continuous, is bounded on compact sets. Then the chain $\{W_n\}$ is ergodic under the same conditions as in Theorem 11.1 except that (i) should be replaced by (i)'. T has a density $g(t)$ which is not concentrated on a bounded set.

PROOF: We find a weakly continuous ϕ-irreducible component of $\{W_n\}$. Our boundedness assumption ensures the existence of a continuous function $k(\omega)$ with $k(\omega) \geqslant \gamma(\omega)$ for every ω. Define transition measures $\tilde{P}(x, \cdot)$ by setting, for each nonnegative measurable f,

$$\tilde{P}(\omega, f) = f(0) \int_{\omega + k(\omega)}^{\infty} g(t)\, dt, \tag{11.7}$$

so that in particular, for any set A,

$$\tilde{P}(\omega, A) = \begin{cases} 0 \text{ if } 0 \notin A \\ \int_{\omega + k(\omega)}^{\infty} g(t)\, dt \text{ if } 0 \in A. \end{cases} \tag{11.8}$$

Since $g(t)$ is not concentrated on any bounded set, (11.8) shows that $\tilde{P}(\omega, \cdot)$ is one-step ϵ_0-irreducible. Moreover, for any f, and in particular for bounded continuous f, (11.7) implies that $\tilde{P}(\omega, f)$ is continuous in ω.

Since $\tilde{P}(\omega, (0, \infty)) = 0$, $P(\omega, A) \geqslant \tilde{P}(\omega, A)$ for any $A \subseteq (0, \infty)$; but also

$$\begin{aligned} P(\omega, 0) &= Pr(T \geqslant \omega + \gamma(\omega) \mid W_n = \omega) \\ &= \int_{\omega+\gamma(\omega)}^{\infty} g(t)\ dt \\ &\geqslant \int_{\omega+k(\omega)}^{\infty} g(t)\ dt \\ &= \tilde{P}(\omega, 0). \end{aligned}$$

Thus $\{\tilde{P}(\omega, \cdot)\}$ is a weakly continuous ϵ_0-irreducible component of $\{P(\omega, \cdot)\}$. The rest of the proof can now be carried out as in Theorem 11.1, but this time $[0, \beta]$ is a test set by virtue of Theorem 5.1 instead of Theorem 4.1.

This theorem covers the type of service times considered by Sugawara and Takahashi [17]; conditions (i) and (i)′ seem to cover the bulk of the likely models for waiting times.

Since, in these two nonoverlapping cases, assumption (ii) gives ergodicity, it may seem that this assumption alone is sufficient for ergodicity, and that steps I and II are unnecessary. The following somewhat artificial example shows that this is not the case; boundedness of the hitting times on compact sets does not give ergodicity if the chain is sufficiently incompatible with the topology of the space.

Example: Suppose the arrivals in the waiting-time-dependent service-time model above are deterministic, with interarrival times of length 1. The waiting times will take on values in $[0, 1]$, and the conditional distribution of the service time S are, for $W \in H = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$,

$$Pr\{S = 1 + (n+1)^{-1} - n^{-1} \mid W = n^{-1}\} = \alpha_n > 0,$$

$$Pr\{S = 2 - n^{-1} \mid W = n^{-1}\} = 1 - \alpha_n > 0,$$

and for $\omega \in [0, 1] \backslash H$,

$$Pr\{S = 1 - \omega \mid W = \omega\} = 1,$$

$$Pr\{S = 2 \mid W = 0\} = 1.$$

Hence $\{W_n\}$ is ϕ-irreducible with ϕ as counting measure on the set H, and in at most two steps from any ω, $\{W_n\}$ takes values in this set ("two-step irreducibility").

Now for all $\omega \geqslant \epsilon$,

$$E[W_1 - W_0 \mid W_0 = \omega] \leqslant -\min\{\epsilon, [\alpha_n(n+1)^{-1} + 1 - \alpha_n - n^{-1}]: n \leqslant \epsilon^{-1}\}.$$

Hence if $\alpha_n \to 1$ sufficiently fast, $\{W_n\}$ will have mean drift towards each set of the form $[0, \epsilon]$. But if $\prod_1^{\infty} \alpha_n > 0$, then the chain has no stationary distribution (intuitively such a distribution would be concentrated at zero, and this cannot happen).

12. DAMS

We consider two generalizations of the Moran model of a dam with infinite capacity (see Prabhu, Ref. [15], Chapter 6). The content Z_n of the dam at time n is determined by a

sequence $\{X_n\}$ of independent identically distributed inputs and the fixed quantity m (= amount released per unit time):

$$Z_{n+1} = Z_n + X_n - \min\{m,\ Z_n + X_n\}. \tag{12.1}$$

We retain the same input scheme but replace the deterministic release rule by a random release rule.

The first model incorporates a "release function" $R(\cdot)$; replace m by $R(Z_n)$ to define

$$\begin{aligned} Z_{n+1} &= Z_n + X_n - \min\{R(Z_n),\ Z_n + X_n\} \\ &= [Z_n + X_n - R(Z_n)]^+; \end{aligned} \tag{12.2}$$

cf: the general release rule for the continuous-time model of Moran [13].

THEOREM 12.1: Suppose the random variables X_n have finite mean $\mu > 0$ and a density function g which is positive at each point of $(0, \infty)$. Then the Markov chain $\{Z_n\}$ is ergodic if

(i) $R(\cdot)$ is continuous,

(ii) $\liminf_{u\to\infty} R(u) > \mu$. (12.3)

PROOF: We shall produce a ϕ-irreducible weakly continuous component of the chain $\{Z_n\}$. This will not only establish the ϕ-irreducibility of the chain (Section 7) but will also prove that bounded sets of positive ϕ-measure are test sets (Theorem 5.1); this takes care of steps I and II of the method for proving ergodicity.

For any nonnegative function f,

$$\begin{aligned} E[f(Z_1)|Z_0 = z] &= \int_0^\infty f[z + x - R(z)]\, g(x)\, dx \\ &\geqslant \int_{[R(z)-z]^+}^\infty f[z + x - R(z)]\, g(x)\, dx \\ &= \tilde{P}(z,\ f),\ \text{say.} \end{aligned}$$

By construction, the substochastic transition laws $\tilde{P}(z, \cdot)$ defined by the above expression constitute a component of (the transition laws of) the chain $\{Z_n\}$.

I. ϕ-irreducibility of the component: Because of (ii) there exists a constant $u_0 > \mu$ such that $R(u) \geqslant \mu + \epsilon$ whenever $u \geqslant u_0$, for some positive ϵ. Let ϕ be Lebesgue measure on $[u_0, \infty]$. Since, by a change of variable,

$$\tilde{P}(z,\ f) = \int_{[z-R(z)]^+}^\infty f(\omega)\, g[\omega + R(z) - z]\, d\omega,$$

and g is everywhere positive on $(0, \infty)$, it follows that $\tilde{P}(z,\ A) > 0$ if $z \leqslant u_0$ and $\phi(A) > 0$. Moreover, since

$$\tilde{P}^{n+1}(z,\ A) \geqslant \int_0^{u_0} \tilde{P}^n(z,\ du)\, \tilde{P}(u,\ A),$$

ϕ-irreducibility will follow if $\tilde{P}^n(z,\ [0,\ u_0]) > 0$ for every $z > u_0$, where n depends upon z. But if $z > u_0$ then $R(z) \geqslant \mu + \epsilon$, and so

$$\begin{aligned} \tilde{P}[z, (z - \mu,\ z - \mu/2)] &\geqslant \int_{z-\mu}^{z-\mu/2} f(\omega)\, g[\omega + R(z) - z]\, d\omega \\ &> 0. \end{aligned}$$

The required conclusion follows by induction.

II. Test sets: We prove that $\tilde{P}(z, \cdot)$ is weakly continuous so that, by Theorem 5.1, any bounded set with positive ϕ-measure will be a test set.

Now if f is bounded and continuous, and if $z_n \to z$, then

$$\tilde{P}(z_n, f) = \int_{[R(z_n)-z_n]^+} f[z_n + x - R(z_n)]\, g(x)\, dx$$
$$\to \int_{[R(z)-z]^+} f[z + x - R(z)]\, g(x)\, dx$$

by dominated convergence since both $(R(z) - z)^+$ and $f(z + x - R(z))$ are continuous functions of z.

III. Boundedness of hitting times: Use Theorem 2.2 with $A = [0, \beta]$ for β large enough.

$$E[Z_1 - Z_0 | Z_0 = z] = E[\max\{X_1 - R(z), z\}]$$
$$\leqslant E[\max\{X_1 - \mu - \epsilon, -z\}] \text{ if } z \geqslant u_0$$
$$\to E[X_1 - \mu - \epsilon] \text{ as } z \to \infty, \text{ by dominated convergence}$$
$$= -\epsilon.$$

We remark that essentially this general release model has been considered in the much more difficult continuous-time case by Cinlar and Pinsky [5] and Harrison and Resnick [8]. Their methods are by necessity much more sophisticated than the straightforward techniques that we can give in the discrete-time case.

The conditions of Theorem 12.1 can be modified slightly to give other conditions for ergodicity:

(a) It suffices for $R(u)$ to be greater than $\mu + \epsilon$ for *every* $u \geqslant 0$; no conditions are needed for the chain $\{X_n\}$. In this case $\{Z_n\}$ is ϵ_0-irreducible. The standard fixed release rule of (12.1) is covered by such an assumption.

(b) Suppose the common density function g of the X_n's is bounded below by a strictly positive, decreasing, continuous function g', and that $R(\cdot)$ is bounded above by a continuous function $R'(\cdot)$ and is bounded below on $[u_0, \infty]$ by a positive continuous function $R''(\cdot)$. Then the method of proof of Theorem 12.1 goes through using the component

$$P^*(z, f) = \int_{[z-R''(z)]^+}^{\infty} f(\omega)\, g'[\omega + R'(z) - z]\, d\omega;$$

notice that $P^*(z, \cdot) \leqslant \tilde{P}(z, \cdot)$.

Further generalizations of the basic dam model are possible. We leave to the reader a proof of the following random-release result, noting only that the assumptions ensure ϵ_0-irreducibility.

THEOREM 12.2: If $\{Y_n\}$ is a sequence of independent, identically distributed, nonnegative random variables which are independent of $\{X_n\}$, we can define the random-release dam model $\{Z_n\}$ by

$$Z_{n+1} = (Z_n + X_n - Y_n)^+.$$

If $E(X_n) < E(Y_n)$ then $\{Z_n\}$ is ergodic.

13. INVENTORIES

The standard inventory model poses a more difficult problem since the transition laws exhibit distinctly noncontinuous behaviour — even the existence of a weakly continuous component would require intuitively unnatural assumptions. However, the chain can be analyzed by means of the result in Section 6.

Let $\{X_n\}$ be the s-S inventory model (Prabhu, Ref. [15], Chapter 5):

$$X_n = X_{n-1} - \xi_n \text{ if } s < X_{n-1} \leqslant S,$$

$$X_n = S - \xi_n \text{ if } X_{n-1} \leqslant s,$$

where the ξ_n are independent, identically distributed non-negative random variables with common distribution law $F(\cdot)$ and $0 \leqslant s \leqslant S$. The chain $\{X_n\}$ has state space $(-\infty, S]$.

THEOREM 13.1: Suppose $F\{0\} < 1$. Then $\{X_n\}$ is ergodic.

PROOF: *I. ϕ-irreducibility:* Clearly $\phi(\cdot) = P(S, \cdot)$ is a suitable irreducibility measure.

II. Test sets: Let μ be a subinvariant measure for the chain. The subinvariant equation (6.1) can be written as

$$\mu(A) \geqslant \int_{(s,S]} \mu(dy)\, P(y, A) + \mu(-\infty, s]\, P(S, A), \tag{13.1}$$

so if A is any set with $P(S, A) > 0$ and $\mu(A) < \infty$ then $\mu(-\infty, s] < \infty$. By assumption, $Pr\{\xi_n > 0\} > 0$ hence $Pr\{\xi_n \geqslant \epsilon\} > 0$ for some $\epsilon > 0$. Thus, if $\mu(-\infty, s] = 0$ then by taking $A = (-\infty, x]$ in (13.1) we would obtain $\mu(s, s + \epsilon] = 0$. Repeated application of this argument would lead to the conclusion that $\mu(-\infty, S] = 0$, which contradicts the assumed nontriviality of μ. Thus $0 < \mu(-\infty, s] < \infty$, and so by Theorem 6.1, $(-\infty, s]$ is a test set.

III. Boundedness of hitting times: Use the test function $g(x) = (x + D)^+$ where D is any fixed positive quantity. Take $A = (-\infty, s]$ as the test set. Then for $x \in A^c = (s, S]$,

$$\begin{aligned} E[g(X_1)|X_0 = x] &= \int_{[0,\infty)} (x - y + D)^+ F(dv) \\ &= x + D - \int_{[0,\infty)} x + D - (x - y + D)^+ F(dy) \\ &= x + D - \int_{[0,\infty)} \min\{x + D, y\}\, F(dy) \\ &\leqslant x + D - \int_{[0,\infty)} \min\{s + D, y\}\, F(dy). \end{aligned}$$

This last integral is strictly positive and independent of the value of x. Notice also that $E[g(X_1)|X_0 = x] \leqslant S + D$ for every x. Thus Theorem 2.1 can be used to prove that $\{X_n\}$ is ergodic.

By a repetition of the argument for step II it could be shown that $\mu(-\infty, S] < \infty$. Thus the whole state space itself is a test set. For this test set the mean drift condition of Theorem 2.2 is trivially satisfied so ergodicity follows as before. We should remark that finiteness of a subinvariant measure can also be shown to be equivalent to ergodicity, by more direct methods.

14. FEEDBACK MODELS

Let $\{Y_n\}$ and $\{Z_n\}$ be mutually independent sets of independent identically distributed random variables, and let b be a positive constant. A general form of the feedback model $\{X_n\}$ is specified by

$$X_{n+1} = \begin{cases} X_n + Y_n & \text{if } |X_n| \leqslant b, \\ X_n + Y_n - Z_n & \text{if } X_n > b, \text{ and} \\ X_n + Y_n + Z_n & \text{if } X_n < -b. \end{cases} \tag{14.1}$$

This model was introduced by Bellman [1], with the Z_n's constant. Calton and Rogers [4] also analyzed a deterministic form of the model, using results similar to our Theorems 2.1 and 8.2.

The model described by (14.1) is not weakly continuous, and so we look for a weakly continuous component. Let $Q_1(A) = Pr\{Y_n \in A\}$, $Q_2(A) = Pr\{Y_n - Z_n \in A\}$ and $Q_3(A) = Pr\{Y_n + Z_n \in A\}$, and write $\tilde{Q}$ for the largest measure such that $\tilde{Q}(A) \leqslant \min[Q_1(A), Q_2(A), Q_3(A)]$ for every Borel set A. A sufficient condition for $\tilde{Q}$ to be nonzero is that Y_n admits a density $g(y)$ satisfying, on some set of positive Lebesgue measure,

$$\min[g(y), \int_{-\infty}^{\infty} g(y+\alpha)\, dPr\{Z_n \leqslant \alpha\}, \int_{-\infty}^{\infty} g(y-\alpha)\, dPr\{Z_n \leqslant \alpha\}] > 0.$$

In the deterministic model, for example, this will occur if

$$\min[g(y), g(y-k), g(y+k)] > 0$$

where the Z_n's take the constant value k; for this it would suffice to have $g(y) > 0$ on some set of the form $(-k-\delta, k+\delta)$.

THEOREM 14.1: Suppose $\tilde{Q}$ dominates Lebesgue measure on some interval $(-\eta, \eta)$. If $|E(Y_n)| < E(Z_n)$ then $\{X_n\}$ is ergodic, whilst if $|E(Y_n)| \leqslant E(Z_n)$ then $\{X_n\}$ is recurrent.

PROOF: Define $\tilde{P}(x, A) = \tilde{Q}(A - x)$. Then by definition $\{\tilde{P}(x, \cdot)\}$ is a component of $\{P(x, \cdot)\}$, and it is ϕ-irreducible with ϕ as Lebesgue measure, from our first assumption (which is again "grossly sufficient"; see Revuz, Ref. [16], Section 3.4, for less practical but finer conditions). Moreover, $\tilde{P}(x, \cdot)$ is weakly continuous: if f is bounded and continuous and $x_n \to x$,

$$\begin{aligned} \tilde{P}(x_n, f) &= \int f(x_n + y)\, \tilde{Q}(dy) \\ &\to \int f(x + y)\, \tilde{Q}(dy) \quad \text{by dominated convergence} \\ &= \tilde{P}(x, f). \end{aligned}$$

It is simple to show that the second assumption implies that $E[|X_1||X_0 = x] \leqslant x - \epsilon$ for all x outside some bounded region $(-M, M)$, whilst $E[|X_1||X_0 = x] \leqslant M + E|Y_0| + E|Z_0| < \infty$ if $x \in (-M, M)$. Ergodicity follows. Similarly the third assumption enables us to use Theorem 8.2, whence the recurrence result.

Using techniques similar to those of Section 12, we could handle feedback models with more general feedback, e.g., we could replace $\{Z_n\}$ in (14.1) by $R(X_n)$. A typical set of conditions for ergodicity would then be that $R(\cdot)$ is continuous (to establish weak continuity), that Y_n has positive density everywhere (to give ϕ-irreducibility with ϕ as Lebesgue measure) and that $\liminf_{x \to \infty} R(x) > E(Y_n)$ and $\limsup_{x \to -\infty} R(x) < E(Y_n)$ (to establish mean drift towards some bounded interval).

REFERENCES

[1] Bellman, R., "Research Problem", Bulletin of the American Mathematical Society, *68*, p. 180 (1962).

[2] Billingsley, P., *Convergence of Probability Measures*, John Wiley, New York, (1968).

[3] Callahan, J. R., "A Queue with Waiting Time Dependent Service Times," Naval Research Logistics Quarterly, *20*, 321-324 (1973).

[4] Calton, W. G., and G. S. Rogers, "On Classifying Discrete Time Markov Processes," submitted for publication (1976).

[5] Cinlar, E., and M. Pinsky, "A Stochastic Integral in Storage Theory," Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete, *17*, 227-240, (1971).

[6] Foster, F. G., "On the Stochastic Process Associated with Certain Queueing Processes," Annals of Mathematical Statistics, *24*, 355-360, (1953).

[7] Feller, W., *An Introduction to Probability Theory and Its Applications,* Vol. I, 3rd ed. (John Wiley, New York, 1970).

[8] Harrison, J. M., and S. I. Resnick, "The Stationary Distribution and First Exit Probabilities of a Storage Process with General Release Rule." submitted for publication (1976).

[9] Kendall, D. G., "Some Problems in the Theory of Queues," Journal of the Royal Statistical Society, B *13*, 151-185 (1951).

[10] Kiefer, J. and J. Wolfowitz, "On the Theory of Queues with Many Servers," Transactions of the American Mathematical Society, *78*, 1-18, (1955).

[11] Loynes, R. M., "The Stability of a Queue with Non-independent Inter-arrival and Service Times," Proceedings of the Cambridge Philosophical Society, *58*, 497-520 (1962).

[12] Mauldon, J. G., "On Non-dissipative Markov chains," Proceedings of the Cambridge Philosophical Society, *53*, 825-835 (1958).

[13] Moran, P. A. P., "A Theory of Dams with Continuous Input and a General Release Rule," Journal of Applied Probability, *6*, 88-98 (1969).

[14] Pakes, A. G., "Some Conditions for Ergodicity and Recurrence of Markov Chains," Operations Research, *17*, 1058-1061 (1967).

[15] Prabhu, N. U., *Queues and Inventories*, (John Wiley, New York, 1965).

[16] Revuz, D., *Markov Chains* (North-Holland, Amsterdam, 1975).

[17] Sugawara, S., and M. Takahashi, "On Some Queues Occurring in an Integrated Iron and Steel Works," Journal of the Operations Research Society of Japan, *8*, 16-23 (1965).

[18] Tweedie, R. L., "R-Theory for Markov Chains on a General State Space I," Annals of Probability, *2*, 840-864, (1974).

[19] Tweedie, R. L., "Criteria for Classifying General Markov Chains," Advances in Applied Probability, *8*, 737-771 (1976).

[20] Tweedie, R.L., "Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on a General State Space," Stochastic Processes and Their Applications, 3, 385-403, (1975).

[21] Tweedie, R.L., "Hitting Times of Markov Chains, with Application to State-Dependent Queues," Bulletin of the Australian Mathematical Society 17, 97-107, (1977).

MAXIMIZING THE SUM OF CERTAIN QUASICONCAVE FUNCTIONS USING GENERALIZED BENDERS DECOMPOSITION

A. Victor Cabot

Graduate School of Business
Indiana University
Bloomington, Indiana

ABSTRACT

In this paper we consider the problem of maximizing the sum of certain quasi-concave functions over a convex set. The functions considered belong to the classes of functions which are known as nonlinear fractional and bi-nonlinear functions. Each individual function is quasi-concave but the sum is not. We show that this nonconvex programming problem can be solved using Generalized Benders Decomposition as developed by Geoffrion.

INTRODUCTION

In this paper we consider the application of the Generalized Benders Decomposition (GBD) as developed by Geoffrion [4] to certain nonconvex programming problems. Before showing how this is done, however, we briefly summarize the application of GBD to a problem whose form is similar to the one of interest.

The variables of our problem are of two forms, $y \in E^n$ and $x \in E^m$. The problem we wish to solve is given by problem P :

P: maximize $z = f(x)$

subject to $G(y,x) \geqslant 0$,

$$x \in X, y \in Y.$$

Of interest here are problems where the functions $G(y,x)$ are not concave in x and y jointly, but fixing x renders them so in y and fixing x yields the same for y. The sets X and Y are taken to be convex sets. The function $f(x)$ is assumed concave.

Goeffrion showed that under certain conditions problem P could be solved for a global maximum, even though it appears at first glance to be a nonconvex programming problem. To achieve this result, Goeffrion extended the work of Benders [2] to include nonlinear programming problems. We now describe the (GBD) method for solving problem P. A knowledge of the method is not necessary but may be helpful. The procedure is represented by the following steps:

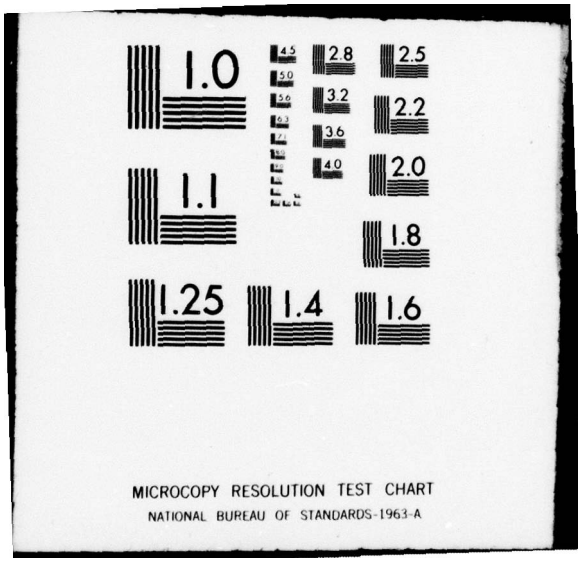
STEP 1: Let a point $\bar{y} \in Y$ be known. Solve the following optimization problem:

P1: maximize $z = f(x)$

AD-A064 991 OFFICE OF NAVAL RESEARCH ARLINGTON VA F/G 15/5
NAVAL RESEARCH LOGISTICS QUARTERLY, VOLUME 25, NUMBER 3,(U)
SEP 78

UNCLASSIFIED NL

2 OF 3
AD
A064991


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

subject to $G(\bar{y},x) \geqslant 0$,

$$x \in X.$$

The value of z in the solution to P1 is a lower bound on the optimal solution to problem P; call it LB. In subsequent steps it may be necessary to solve P1 several times. The value of LB is always given as the greatest of the values of z in solutions to problem P1. Denote by $\bar{u} \geqslant 0$ Kuhn-Tucker multipliers associated with the constraints $G(y,x) \geqslant 0$ in the solution to problem P1. We assume that such multipliers are given by the method used for solving problem P1. We also assume that for any $\bar{y} \in Y$ there exists an $x^{\cdot} \in X$ such that $G(\bar{y},x^{\cdot}) \geqslant 0$. Thus, problem P1 will always have a feasible solution. Now go to Step 2 with $p=1$ and $u^1=\bar{u}$.

STEP 2: Solve the optimization problem

P1A: maximize y_o

subject to $y_o \leqslant \underset{x \in X}{\text{maximum}} \{f(x) + u^j g(y,x)\}$, $j = 1, \ldots, p$,

$$y \in Y.$$

Let $(\hat{y}_o,\hat{y})$ solve problem P1A. Let $UB = \hat{y}_o$, where UB is now a known upper bound on the value of the optimal solution to problem P.

STEP 3: Return to problem P1 with the solution $\hat{y}$ and solve it. If $LB \geqslant UB - \epsilon$ (where ϵ is a prechosen number), terminate with $\hat{y}$ and the corresponding solution to problem P1 as the solution to problem P. Otherwise, increase p by one and put $u^p = \bar{u}$ and return to Step 2. We now point out various aspects of the algorithm.

Note that in Step 1 the optimization problem is essentially problem P with the vector y fixed. The lower bound created at this stage may not increase at each iteration of the algorithm. We assume that problem P1 satisfies some sort of regularity assumption so that the vector $\bar{u}$ exists.

Step 2 of the method is the most difficult to perform. Problem P1A is essentially the dual of problem P. The value of y_o in the solution to problem P1A will be strictly decreasing, since each problem is more constrained than its predecessor. We thus see that the value of UB will decrease at each iteration until the method terminates. The solution of problem P1A will generally be in two steps. First, compute the maximum on the right-hand side of the constraints corresponding to u^p. Second, solve problem P1A for the values of $(\hat{y}_o,\hat{y})$. The first of these steps is the most difficult, unless the maximum can be carried out independently of the vector y. Note that once this maximization is carried out, the second step is straightforward if the function $G(y,x)$ is assumed to be concave in y for fixed x.

Termination of the method in a finite number of steps (for $\epsilon > 0$) is provided by Geoffrion in the following theorm, modified slightly to correspond to problem P.

THEOREM (Geoffrion, Ref. [4]): Assume that X and Y are compact convex sets and that for every $\bar{y} \in Y$ there is an $x^* \in X$ such that $G(\bar{y},x^*) \geqslant 0$. Also, assume that the functions f and G are concave on X for every fixed $y \in Y$, and that problem P1 satisfies a regularity assumption guaranteeing multipliers $\bar{u}$ for every $y \in Y$. Then, for any $\epsilon > 0$, the *GBD* procedure terminates in a finite number of steps.

In the following sections, we shall show that the assumptions of the above theorem are satisfied by two problems which are equivalent to nonconvex programming problems.

NONLINEAR FRACTIONAL PROGRAMMING

In this section we consider a special form of problem P. While the constraint set is a convex set, the objective function is not concave, and thus maximization may lead to locally optimal but globally suboptimal solutions. The problem takes the form:

$$\text{P: maximize} \quad z = \sum_{i=1}^{m} f_i(y)/c_i(y)$$

subject to $y \in Y$

Once again we assume that $y \in E^n$ and that Y is a convex set. Each term of the objective function is the ratio of $f_i(y)$, a concave function, to $c_i(y)$, a linear function in y. We assume that $c_i(y) > 0$ and $f_i(y) \geq 0$ for $y \in Y$, $i=1, \ldots, m$. Such functions belong to the class of functions which Mangasarian [5] terms nonlinear fractional functions.

It is well known that each function $f_i(y)/c_i(y)$ is a quasi-concave function, in the sense that for constant $x_i \geq 0$, the set $f_i(y)/c_i(y) \geq x_i$ is a convex set. We shall use this result to formulate another optimization problem which is equivalent to problem P, in the sense that both problems have the same optimal solution. We will apply the *GBD* procedure to the equivalent problem.

To formulate the equivalent problem, we make note of the fact that it is always possible to determine a value of x_i^u so large that the inequality has no solution for $y \in Y$. If this is not true, then the problem would have an infinite solution. In practice one could determine x_i^u by maximizing each function $f_i(y)/c_i(y)$ over Y. (Theoretically, one must assume $f_i(y)/c_i(y)$ is pseudoconcave to perform this maximization [1,5], but in actual practice this rarely causes difficulties.) We thus assume that one can obtain the numbers x_i^u, $i=1, \ldots, m$.

Given the above we formulate problem P′ :

$$\text{P′: maximize } z = \sum_{i=1}^{m} x_i$$

subject to $f_i(y) - x_i\, c_i(y) \geq 0$, $i=1 \ldots, m$,

$$y \in Y, 0 \leq x_i \leq x_i^u, i=1, \ldots, m.$$

We see that the optimal solutions to problem P and problem P′ are the same. For notational purposes we denote the set $X = \{x_i | 0 \leq x_i \leq x_i^u \quad i=1, \ldots, m\}$. Comparing the components of problem P′ to the convergence theorem for the *GBD* procedure we find that the sets Y and X are compact, and that for any $\bar{y} \in Y$ there exists $x^* \in X$ such that $G(\bar{y}, x^*) = f_i(\bar{y}) - x_i^* c_i(\bar{y}) \geq 0$, $i=1, \ldots, m$. We note also that $f_i(y) - x_i\, c_i(y)$ is concave in x_i for fixed y. Thus, if we can perform the steps of the method we can guarantee that, for any $\epsilon > 0$, the *GBD* procedure will yield the optimal solution to problem P′ in a finite number of steps.

Applying the *GBD* procedure to problem P′ we must first solve at each stage the problem P′ 1 with y fixed at $\bar{y} \in Y$,

$$\text{P′1 maximize } z = \sum_{i=1}^{m} x_i$$

subject to $f_i(\bar{y})-x_i c_i(\bar{y}) \geq 0, i=1, \ldots, m,$

$$x \in X.$$

It is easily seen that problem P' 1 is in fact equivalent to m linear programming problems in one variable, and thus has the closed form solution $\bar{x}_i = f_i(y)/c_i(\bar{y}), i=1, \ldots, m$. Since P' 1 is a linear-programming problem, the Kuhn-Tucker multipliers always exist and in fact have closed form solution $\bar{u}_i = 1/c_i(\bar{y})$, $i=1, \ldots, m$.

The first time problem P'1A is solved it will have the form

P'1A: maximize y_o

$$\text{subject to } y_o \leq \underset{x \in X}{\text{maximum}} \sum_{i=1}^{m} x_i + \sum_{i=1}^{m} \bar{u}_i \{f_i(y)-x_i c_i(y)\},$$

$$y \in Y.$$

Because of the form of X this yields the problem

P' 1A: maximize y_o

$$\text{subject to } y_o \leq \sum_{i=1}^{m} \bar{u}_i f_i(y) + \sum_{i=1}^{m} \underset{0 \leq x_i \leq x_i^u}{\text{maximum}} \{x_i[1-\bar{u}_i c_i(y)]\},$$

$$y \in Y.$$

For a typical term on the right-hand side of the constraint, the maximization will be given by $x_i = 0$, if $1 - \bar{u}_i c_i(y) \leq 0$, and $x_i = x_i^u$, if $1 - \bar{u}_i c_i(y) \geq 0$. Of course, when $1 - \bar{u}_i c_i(y) = 0$, the value of x_i is of no consequence. Thus, in order to solve problem P'1A, one must solve 2^m problems considering all possible signs of the terms $1 - \bar{u}_i c_i(y)$. For example, when the enumeration is performed one might start with $1 - \bar{u}_i c_i(y) \leq 0$ for all i and solve the problem

P' 1A : maximize y_o

$$\text{subject to } y_o \leq \sum_{i=1}^{m} \bar{u}_i f_i(y),$$

$$1-\bar{u}_i c_i(y) \leq 0, i=1, \ldots, m,$$

$$y \in Y.$$

This problem is a concave programming problem, since $\bar{u}_i \geq 0$, $c_i(y)$ is linear, and $f_i(y)$ is concave for $i=1, \ldots, m$. We might next consider the problem with $1 - \bar{u}_1 c_1(y) \geq 0$ and all other conditions unchanged.

This gives the problem

P' 1A : maximize y_o

subject to $y_o \leqslant \sum_{i=1}^{m} \bar{u}_i f_i(y)[1-\bar{u}_1 c_1(y)]$,

$$1-\bar{u}_1 c_1(y) \geqslant 0, 1-\bar{u}_i c_i(y) \leqslant 0, i=2, \ldots, m, y \in Y.$$

This is also a concave program due to the linearity of $c_1(y)$.

In order to solve problem P'1A one must first solve all 2^m cases and take as $(\hat{y}_o, \hat{y})$ the solution with the largest value of $\hat{y}_o$. This becomes an upper bound on the optimal solution to problem P'. Of course, this will be quite cumbersome if m is large, but many practical problems are still within the range of solvability since it is the number of quasi-concave functions in the sum, rather than the number of variables in the problem, that determines the degree of difficulty of the problem.

There are two additional points of difficulty. Once problem P'1A has been solved, it is necessary to carry the contraints on the quantities $1 - \bar{u}_i c_i(y)$ forward to the next formulation of P'1A. Thus, at each step of the method the constraints on P'1A grow larger. Since they are linear, this may not increase the difficulty of P'1A too much. A more serious problem occurs when two or more different solutions at any one stage tie for having the maximum value of $\hat{y}_o$. In this case, it is necessary to carry all these solutions to the next stage, solve problem P' 1 for each $\bar{y}$ (unless they are the same) and then solve P' 1A for each combination. Thus, if k solutions tie in the solution to problem P'1A at any stage, it might be necessary to solve, at a minimum, $k2^m$ programming problems the next time problem P'1A occurred.

The combinatorial nature of the solution procedure may at first seem depressing. It should be pointed out, however, that it has been shown [3] that zero-one integer programming problems can be formulated as a special case of this problem. Viewed from this perspective, it is not surprising that the effectiveness of the procedure deteriorates exponentially in the number of functions in the sum. Keeping this in mind, we present an example of the method in the next section.

AN EXAMPLE PROBLEM

Consider the following problem:

P: maximize $z = y_1+1/y_1+y_2 + y_2/3-y_2$

subject to $y \in Y = \{y_1, y_2 | y_1+y_2 \geqslant 1,$

$$y_1+y_2 \leqslant 2, y_1 \geqslant 0, y_2 \geqslant 0\}.$$

A geometric view of the problem with the value of z given at each extreme point appears in Figure 1. It appears, from the extreme points, that there are two local maxima. It should be pointed out, however, that there is no guarantee that the global maximum occurs at an extreme point. Taking $x_1^u = x_2^u = 2$, we formulate problem P' as

P': maximize $z = x_1+x_2$

subject to $y_1+1-x_1(y_1+y_2) \geqslant 0,$

$$y_2-x_2(3-y_2) \geqslant 0,$$

$$y \in Y = \{y_1, y_2 | y_1+y_2 \geqslant 1, y_1+y_2 \leqslant 2, y_1 \geqslant 0, y_2 \geqslant 0\},$$

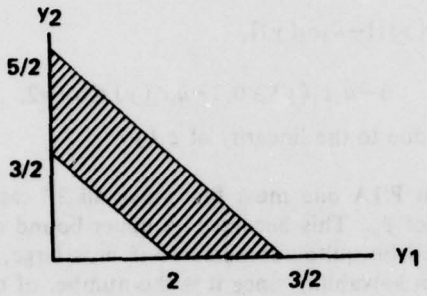


FIGURE 1

and

$$x \in X = \{x_1, x_2 | 0 \leqslant x_1 \leqslant 2, 0 \leqslant x_2 \leqslant 2\}.$$

To begin, we pick the local optimum, $\bar{y}_1 = 1$, $\bar{y}_2 = 0$, and get problem P' 1:

P' 1: maximize $z = x_1 + x_2$

subject to $x_1 \leqslant 2, 3x_2 \leqslant 0, x \in X$, which yields $x_1 = 2$, $x_2 = 0$, $z = 2$, so that $LB = 2$. We get multipliers $\bar{u}_1 = 1$, $\bar{u}_2 = 1/3$. Problem P'1A has the form:

P'1A: maximize y_o

subject to $y_o \leqslant \underset{\substack{0 \leqslant x_1 \leqslant 2 \\ 0 \leqslant x_2 \leqslant 2}}{\text{maximum}} \{x_1 + x_2 + [y_1 + 1 - x_1(y_1 + y_2)] + 1/3[y_2 - x_2(3 - y_2)]\}$,

$$y \in Y.$$

Simplifying, we get the problem

P'1A: maximize y_o

subject to $y_o \leqslant -y_1$

$$+(1/3)y_2 + 1 + \underset{\substack{0 \leqslant x_1 \leqslant 2 \\ 0 \leqslant x_2 \leqslant 2}}{\text{maximum}} \{x_1[1 - (y_1 + y_2)] + x_2[1 - (1/3)(3 - y_2)]\},$$

$$y \in Y.$$

We must solve four problems:

1: maximize y_o

subject to $y_o \leqslant -y_1 - y_2 + 3$,

$$y_1 + y_2 \leqslant 1, (1\, y_2 \geqslant 0, y \in Y,$$

with optimal solution $y_1 = 1, y_2 = 0, y_o = 2$.

2: maximize y_o

subject to $y_o \leqslant -y_1-(5/3)y_2+3$,

$$y_1+y_2 \leqslant 1,\ (1/3)\,y_2 \leqslant 0, y \in Y,$$

with optimal soluton $y_1 = 1,\ y_2 = 0,\ y_o = 2$.

3: maximize y_o

subject to $y_o \leqslant y_1 + y_2 + 1$,

$$y_1 + y_2 \geqslant 1,\ (1/3)y_2 \geqslant 0, y \in Y,$$

with optimal solution $y_1 = 2, y_2 = 0, y_o = 3$.

4: maximize y_o

subject to $y_o \leqslant y_1+(1/3)y_2+1$,

$$y_1+y_2 \geqslant 1,\ (1/3)y_2 \leqslant 0, y \in Y,$$

with optimal solution $y_1 = 2, y_2 = 0, y_o = 3$.

Problems 3 and 4 have the same value of y_o (we take this solution to problem 3 to aid the explanation, although the solution $y_1 = 0,\ y_2 = 2, y_0 = 3$ has the same value), and thus $y_o = 3$ serves as an upper bound on the value of z in problem P'. We thus proceed to problem $P'1$ with $LB = 2$, $UB = 3$, and the solution $\bar{y}_1 = 2, \bar{y}_2 = 0$.

P'1: maximize $z = x_1+x_2$

subject to $2x_1 \leqslant 3,\ 3x_2 \leqslant 0,\ x \in X$. The solution gives $x_1 = 3/2,\ x_2 = 0$, and $z = 3/2$ so that $LB = 2$ again and $\bar{u}_1 = 1/2, \bar{u}_2 = 1/3$.

Proceeding to problem P'1A, we get the cases:

1: maximize y_o
subject to $y_o \leqslant y_1+(1/3)y_2+1$,
$y_o \leqslant (1/2)y_1+(1/3)y_2+1/2$,
$y_1+y_2 \geqslant 1$,
$y_1+y_2 \geqslant 2$,
$(1/3)y_2 \geqslant 0$,
$y \in Y$,
Solutions: $y_1 = 2, y_2 = 0,\ y_o = 3/2$.

2: maximize y_o
subject to $y_o \leqslant y_1+(1/3)y_2+1$,
$y_o \leqslant (1/2)y_1+y_2+1/2$,
$y_1+y_2 \geqslant 1$,
$y_1+y_2 \geqslant 2$,
$(1/3)y_2 = 0$,
$y \in Y$.

maximize y_o
subject to $y_o \leqslant y_1+1$,
$y_o \leqslant (1/2)y_1+(1/3)y_2+1/2$,
$y_1+y_2 \geqslant 1$,
$y_1+y_2 \geqslant 2$,
$(1/3)y_2 = 0$,
$y \in Y$.
Solutions: $y_1 = 1, y_2 = 0,\ y_o = 3/2$.

maximize y_o
subject to $y_0 \leqslant y_1+y_2+1$,
$y_o \leqslant (1/2)y_1+(1/3)y_2+1/2$,
$y_1+y_2 \geqslant 1$,
$y_1+y_2 \geqslant 2$,
$(1/3)y_2 \geqslant 0$,
$y \in Y$.

Solutions: $y_1 = 2, y_o = 2, y_2 = 0,\ y_o = 3/2.$

3: maximize y_o
subject to $y_o \leqslant y_1 + (1/3)y_2 + 1,$
$y_o \leqslant (1/2)y_1 - (2/3)y_2 + 5/2,$
$y_1 + y_2 \geqslant 1,$
$y_1 + y_2 \leqslant 2,$
$(1/3)y_2 \leqslant 0,$
$y \in Y.$

Solutions: $y_1 = 1, y_2 = 0,\ y_o = 2.$

4: maximize y_o
subject to $y_o \leqslant y_1 + (1/3)y_2 + 1,$
$y_o \leqslant -(1/2)y_1 + 5/2,$
$y_1 + y_2 \leqslant 1,$
$y_1 + y_2 \leqslant 2,$
$(1/3)y_2 = 0,$
$y \in Y.$

Solutions: $y_1 = 1, y_2 = 0,\ y_o = 2.$

Solutions: $y_1 = 0,\ y_2 = 2, y_o = 5/2.$

maximize y_o
subject to $y_o \leqslant y_1 + y_2 + 1,$
$y_o \leqslant -(1/2)y_1 - (2/3)y_2 + 5/2,$
$y_1 + y_2 \geqslant 1,$
$y_1 + y_2 \leqslant 2,$
$(1/3)y_2 = 0,$
$y \in Y.$

Solutions: $y_1 = 1, y_2 = 0,\ y_o = 2.$

maximize y_o
subject to $y_o \leqslant y_1 + y_2 + 1,$
$y_o \leqslant -(1/2)y_1 + 5/2,$
$y_1 + y_2 \geqslant 1,$
$y_1 + y_2 \leqslant 2,$
$(1/3)y_2 \geqslant 0,$
$y \in Y.$

Solutions: $y_1 = 0,\ y_2 = 2, y_o = 5/2.$

The best solution is $y_1 = 0,\ y_2 = 2$, and $y_o = 5/2$, which gives $UB = 5/2$. When the solution $y_1 = 0,\ y_2 = 2$ is used in problem P'1, we get $LB = 5/2$, and thus the optimal solution appears to be the extreme point corresponding to the largest value of z.

In the following section we consider another class of quasi-concave function for which the *GBD* procedure is applicable.

BI-NONLINEAR PROGRAMMING

A second type of quasiconcave function amenable to the proposed technique is a bi-nonlinear function. The proposed problem is of the following kind:

$$\text{P: maximize } z = \sum_{i=1}^{m} f_i(y) \cdot c_i(y)$$

subject to $y \in Y$. The functions $f_i(y)$ are assumed to be concave, and the functions $c_i(y)$ to be linear. We also assume that $f_i(y) \geqslant 0$ and $c_i(y) > 0$ for $y \in Y,\ i = 1, \ldots, m$ [5]. The problem P′ is formed as follows:

$$\text{P': maximize } z = \sum_{i=1}^{m} x_i$$

subject to $f_i(y) - x_i/c_i(y) \geqslant 0,\ i = 1, \ldots, m, \quad y \in Y, \quad x \in X.$

The set X is determined exactly as it was previously. Note that this problem also satisfies the assumptions for solution by the *GBD* procedure. Once again, for fixed y problem P′ has a closed form solution as m linear programming problems, so that Kuhn-Tucker multipliers u_i are again guaranteed to exist.

For a given u vector, problem P'1A can be formulated as follows:

P'1A: maximize y_o

$$\text{subject to } y_o \leqslant \sum_{i=1}^{m} u_i f_i(y) + \sum_{i=1}^{m} \underset{0 \leqslant x_i \leqslant x_i^u}{\text{maximum}} \{x_i[1-u_i/c_i(y)]\},$$

$$y \in Y.$$

Here, once again, the solution in terms of x_i depends on the sign of the terms $1-u_i/c_i(y)$. If $1-u_i/c_i(y) \leqslant 0$, we have $x_i = 0$, and if $1-u_i/c_i(y) \geqslant 0$, we have $x_i = x_i^u$. Thus, the solution procedure for this problem exactly parallels the procedure for the nonlinear fractional problem. It should be pointed out, however, that problem P'1A is not a sequence of linear programs, since it contains the terms $u_i/c_i(y)$ for problems with $x_i = x_i^u$. A saving grace may be that since the functions $c_i(y)$ are linear, a simple change of variables can transform these terms into separable nonlinear functions for which linear-programming approximations are available.

REFERENCES

[1] Arrow, K., and A. Enthoven, "Quasi-Concave Programming," Econometrica, *29* (1961).
[2] Benders, J.F., "Partitioning Procedures for Solving Mixed-Variables Programming Problems," Numerische Mathematik, *4* (1962).
[3] Cabot, A.V., "On the Generalized Lattice Point Problem and Nonlinear Programming," Operations Research *23*, (1975).
[4] Geoffrion, A.M., "Generalized Benders Decomposition," Journal of Optimization Theory and Applications, *10*, (1972).
[5] Mangasarian, O.L., Nonlinear Programming, McGraw Hill Book Co., New York, (1969).

A HETEROGENEOUS ARRIVAL AND SERVICE QUEUEING LOSS MODEL*

Simson Fond

Department of Mathematics
George Mason University
Fairfax, Virginia

Sheldon M. Ross

Department of Industrial Engineering
and Operations Research
University of California, Berkeley
Berkeley, California

ABSTRACT

Consider a single-server exponential queueing loss system in which the arrival and service rates alternate between the pairs (λ_1, μ_1) and (λ_2, μ_2), spending an exponential amount of time with rate $c\alpha_i$ in (λ_i, μ_i), $i = 1, 2$. It is shown that if all arrivals finding the server busy are lost, then the percentage of arrivals lost is a decreasing function of c. This is in line with a general conjecture of Ross to the effect that the "more nonstationary" a Poisson arrival process is, the greater the average customer delay (in infinite capacity models) or the greater the percentage of lost customers (in finite capacity models). We also study the limiting cases when c approaches 0 or infinity.

1. INTRODUCTION

This paper is a continuation of a study of queueing models with nonstationary Poisson arrivals begun in Ref. [2], where it was conjectured, and verified in a special case, that a queueing system with nonstationary Poisson arrivals will lead to larger average customer delays than would a similar model having stationary Poisson arrivals with the same average arrival rate. In order to investigate this conjecture further we consider a single-server loss system that oscillates between two feasible levels denoted by 1 and 2. When the system is at level i ($i = 1,2$), the arrival process is a Poisson process with rate λ_i and the service times are exponential random variables with rate μ_i. The time interval during which the system functions at level i is also an exponential random variable with rate $c\alpha_i$ where c is a constant, i.e. the persistence of the system at any level is governed by a random mechanism: if the system is functioning at level i, it tends to "jump" to the alternative level with Poisson rate $c\alpha_i$.

*This research has been partially supported by the Office of Naval Research under Contract N00014-77-C-0299 and the Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-77-3213 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

We suppose that an arriving customer will only enter the system if the server is free when he arrives. Let $L(c)$ denote the proportion of customers that are lost to the system. In the following section we show that

$$L(c) \text{ is decreasing and convex in } c.$$

It should be noted that the (time) average arrival and service rates, $\bar{\lambda}$ and $\bar{\mu}$, are given by

$$\bar{\lambda} = \frac{\lambda_1\alpha_2 + \lambda_2\alpha_1}{\alpha_1 + \alpha_2}, \qquad \bar{\mu} = \frac{\mu_1\alpha_2 + \mu_2\alpha_1}{\alpha_1 + \alpha_2}$$

and are thus independent of c. The purpose of the constant c is to regulate how fast the system changes levels; thus, the larger c is, in some sense "the more stationary the process is." Indeed, as c approaches infinity, the system converges to a stationary one.

2. THE LOSS FUNCTION L(c)

The system can be analyzed as a continuous-time Markov process with states $\{(m,i) | m = 0,1 \text{ and } i = 1,2\}$, where m denotes the number of customers in the system and i denotes the level of the system. The transition probabilities are stationary and satisfy the forward Kolmogorov differential equations. Moreover, for all (m,i), the limiting probabilities, call them P_{mi}, exist and are independent of the initial state. The set $\{P_{mi}\}$ satisfies the following balance equations:

$$(\lambda_1 + c\alpha_1)P_{01} = \mu_1 P_{11} + c\alpha_2 P_{02} \tag{1a}$$

$$(\mu_1 + c\alpha_1)P_{11} = \lambda_1 P_{01} + c\alpha_2 P_{12} \tag{1b}$$

$$(\lambda_2 + c\alpha_2)P_{02} = \mu_2 P_{12} + c\alpha_1 P_{01} \tag{2a}$$

$$(\mu_2 + c\alpha_2)P_{12} = \lambda_2 P_{02} + c\alpha_1 P_{11} \tag{2b}$$

with

$$P_{01} + P_{11} + P_{02} + P_{12} = 1. \tag{3}$$

Let $L(c)$ denote the proportion of customers lost to the system. Since

$$\bar{\lambda}L(c) = \lambda_1 P_{11} + \lambda_2 P_{12},$$

we can calculate $L(c)$ by finding P_{11} and P_{12}. Before doing that, let us note that the proportion of time the system is in level 1 is

$$P_{01} + P_{11} = \frac{\alpha_2}{\alpha_1 + \alpha_2} \tag{4}$$

which can be obtained either by adding (1a) and (1b) together and substituting (3), or by considering the system as an alternating renewal process. Similarly,

$$P_{02} + P_{12} = \frac{\alpha_1}{\alpha_1 + \alpha_2}. \tag{5}$$

The easiest way to solve for P_{11} and P_{12} is to put (1a) and (1b) in a matrix form as follows:

$$\begin{bmatrix} (\lambda_1 + c\alpha_1) & -\mu_1 \\ -\lambda_1 & (\mu_1 + c\alpha_1) \end{bmatrix} \begin{bmatrix} P_{01} \\ P_{11} \end{bmatrix} = \begin{bmatrix} c\alpha_2 P_{02} \\ c\alpha_2 P_{12} \end{bmatrix}. \tag{6}$$

Similarly, for (2a) and (2b):

$$\begin{bmatrix} (\lambda_2 + c\alpha_2) & -\mu_2 \\ -\lambda_2 & (\mu_2 + c\alpha_2) \end{bmatrix} \begin{bmatrix} P_{02} \\ P_{12} \end{bmatrix} = \begin{bmatrix} c\alpha_1 P_{01} \\ c\alpha_1 P_{11} \end{bmatrix}. \tag{7}$$

Putting (6) and (7) together yields

$$\begin{bmatrix} (\lambda_1 + c\alpha_1) & -\mu_1 \\ -\lambda_1 & (\mu_1 + c\alpha_1) \end{bmatrix} \begin{bmatrix} (\lambda_2 + c\alpha_2) & -\mu_2 \\ -\lambda_2 & (\mu_2 + c\alpha_2) \end{bmatrix} \begin{bmatrix} P_{02} \\ P_{12} \end{bmatrix} = \begin{bmatrix} c^2\alpha_1\alpha_2 P_{02} \\ c^2\alpha_1\alpha_2 P_{12} \end{bmatrix}. \tag{8}$$

From the first row of (8) we obtain

$$[c(\alpha_1\lambda_2 + \alpha_2\lambda_1) + \lambda_2(\lambda_1 + \mu_1)]P_{02} = [c(\alpha_1\mu_2 + \alpha_2\mu_1) + \mu_2(\lambda_1 + \mu_1)]P_{12}.$$

Therefore,

$$\frac{P_{12}}{P_{02}} = \frac{c(\alpha_1\lambda_2 + \alpha_2\lambda_1) + \lambda_2(\lambda_1 + \mu_1)}{c(\alpha_1\mu_2 + \alpha_2\mu_1) + \mu_2(\lambda_1 + \mu_1)}. \tag{9}$$

Hence, by (5) and (9),

$$P_{12} = \frac{\alpha_1}{\alpha_1 + \alpha_2} \cdot \frac{c(\alpha_1\lambda_2 + \alpha_2\lambda_1) + \lambda_2(\lambda_1 + \mu_1)}{c[\alpha_1(\lambda_2 + \mu_2) + \alpha_2(\lambda_1 + \mu_1)] + (\lambda_1 + \mu_1)(\lambda_2 + \mu_2)}$$

and

$$P_{02} = \frac{\alpha_1}{\alpha_1 + \alpha_2} \cdot \frac{c(\alpha_1\mu_2 + \alpha_2\mu_1) + \mu_2(\lambda_1 + \mu_1)}{c[\alpha_1(\lambda_2 + \mu_2) + \alpha_2(\lambda_1 + \mu_1)] + (\lambda_1 + \mu_1)(\lambda_2 + \mu_2)}.$$

Due to the symmetry of the equations (1a), (1b) and (2a), (2b), we see that

$$P_{11} = \frac{\alpha_2}{\alpha_1 + \alpha_2} \cdot \frac{c(\alpha_1\lambda_2 + \alpha_2\lambda_1) + \lambda_1(\lambda_2 + \mu_2)}{c[\alpha_1(\lambda_2 + \mu_2) + \alpha_2(\lambda_1 + \mu_1)] + (\lambda_1 + \mu_1)(\lambda_2 + \mu_2)}$$

$$P_{01} = \frac{\alpha_2}{\alpha_1 + \alpha_2} \cdot \frac{c(\alpha_1\mu_2 + \alpha_2\mu_1) + \mu_1(\lambda_2 + \mu_2)}{c[\alpha_1(\lambda_2 + \mu_2) + \alpha_2(\lambda_1 + \mu_1)] + (\lambda_1 + \mu_1)(\lambda_2 + \mu_2)}.$$

Thus, we have

$$\bar{\lambda}L(c) = \frac{c(\alpha_1\lambda_2 + \alpha_2\lambda_1)^2 + \lambda_2[\alpha_2(\lambda_2 + \mu_2) + \lambda_2(\alpha_1(\lambda_1 + \mu_1)}{(\alpha_1 + \alpha_2)[c[\alpha_1(\lambda_2 + \mu_2) + \alpha_2(\lambda_1 + \mu_1)] + (\lambda_1 + \mu_1)(\lambda_2 + \mu_2)]}.$$

Differentiation yields

$$\bar{\lambda}L'(c) = \frac{-\alpha_1\alpha_2(\lambda_1\mu_2 - \lambda_2\mu_1)^2}{(\alpha_1 + \alpha_2)[c[\alpha_1(\lambda_2 + \mu_2) + \alpha_2(\lambda_1 + \mu_1)] + (\lambda_1 + \mu_1)(\lambda_2 + \mu_2)]^2}$$

and

$$\bar{\lambda}L''(c) = \frac{2\alpha_1\alpha_2(\lambda_1\mu_2 - \lambda_2\mu_1)^2[\alpha_1(\lambda_2 + \mu_2) + \alpha_2(\lambda_1 + \mu_1)]}{(\alpha_1 + \alpha_2)[c[\alpha_1(\lambda_2 + \mu_2) + \alpha_2(\lambda_1 + \mu_1)] + (\lambda_1 + \mu_1)(\lambda_2 + \mu_2)]^3}.$$

There are two cases to consider.

CASE 1: $\lambda_1\mu_2 - \lambda_2\mu_1 = 0$; i.e., the traffic intensities λ_1/μ_1 and λ_2/μ_2 are equal, say to ρ.

In this case $L'(c) = 0$, and thus $L(c)$ is independent of the value c. Moreover, we have simple solutions for the P_{mi}'s in this case, namely

$$P_{01} = \frac{\alpha_2}{\alpha_1 + \alpha_2} \frac{1}{1+\rho}$$

$$P_{11} = \frac{\alpha_2}{\alpha_1 + \alpha_2} \frac{\rho}{1+\rho}$$

$$P_{02} = \frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{1}{1+\rho}$$

$$P_{12} = \frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{\rho}{1+\rho}.$$

Hence, P_1, the proportion of time the system is busy, is

$$P_1 \equiv P_{11} + P_{12} = \frac{\rho}{1+\rho},$$

and P_0, the proportion of time the system is empty, is

$$P_0 \equiv P_{01} + P_{02} = \frac{1}{1+\rho}.$$

In terms of P_0 and P_1, the system functions as an ordinary $M/M/1$ loss system with traffic intensity ρ. The loss function is found to be

$$L(c) = \frac{\rho}{1+\rho}.$$

CASE 2: $\lambda_1\mu_2 - \lambda_2\mu_1 \neq 0$.

In this case $L'(c) < 0$ and $L''(c) > 0$. Hence, $L(c)$ is a decreasing convex function of the value c.

Therefore, if the ratio of the time the system stays at each level is fixed, then the faster the system alternates between these two levels, the better the system is (in terms of the loss function).

3. EXTREME CASES

We have shown that $L(c)$ is a strictly decreasing function of c when the traffic intensities λ_1/μ_1 and λ_2/μ_2 are not equal. Now let us study the two extreme cases: (a) $c \to \infty$, i.e. the system alternates extremely fast between level 1 and level 2 or, equivalently, the mean time that the system stays at each level approaches 0; (b) $c \to 0$, i.e. the system alternates extremely slowly between level 1 and level 2 or, equivalently, the mean time that the system stays at each level is becoming infinitely large.

CASE 1: $c \to \infty$

$$\bar{\lambda} \lim_{c\to\infty} L(c) = \frac{(\alpha_1\lambda_2 + \alpha_2\lambda_1)^2}{(\alpha_1 + \alpha_2)[\alpha_1(\lambda_2 + \mu_2) + \alpha_2(\lambda_1 + \mu_1)]},$$

implying that

$$\lim_{c\to\infty} L(c) = \frac{\bar{\lambda}}{\bar{\mu} + \bar{\lambda}}.$$

Furthermore, the proportion of time that the system is busy can be obtained by

$$P_1 = \lim_{c\to\infty} P_{11} + \lim_{c\to\infty} P_{12} = \frac{\bar{\lambda}}{\bar{\mu}+\bar{\lambda}},$$

and the proportion of time that the system is idle is

$$P_0 = \lim P_{01} + \lim P_{02} = \frac{\bar{\lambda}}{\bar{\mu}+\bar{\lambda}}.$$

Thus, the limiting system is equivalent to a no-queue-allowed $M/M/1$ system with constant arrival rate $\bar{\lambda}$ and service rate $\bar{\mu}$.

Since $L(c)$ is decreasing, the value $\dfrac{\bar{\lambda}}{\bar{\mu}+\bar{\lambda}}$ is the smallest value the system can achieve for the loss function.

CASE 2: $c \to 0$

$$\bar{\lambda} \lim L(c) = \frac{\lambda_1^2\alpha_2(\lambda_2+\mu_2) + \lambda_2^2\alpha_1(\lambda_1+\mu_1)}{(\alpha_1+\alpha_2)(\lambda_1+\mu_1)(\lambda_2+\mu_2)}$$

$$= \frac{\alpha_2}{\alpha_1+\alpha_2}\,\frac{\lambda_1^2}{\lambda_1+\mu_1} + \frac{\alpha_1}{\alpha_1+\alpha_2}\,\frac{\lambda_2^2}{\lambda_2+\mu_2},$$

and the proportion of time that the system is busy is

$$P_1 = \lim_{c\to 0} P_{11} + \lim_{c\to 0} P_{12} = \frac{\alpha_2}{\alpha_1+\alpha_2}\,\frac{\lambda_1}{\mu_1+\lambda_1} + \frac{\alpha_1}{\alpha_1+\alpha_2}\,\frac{\lambda_2}{\mu_2+\lambda_2}.$$

The proportion of time the system is idle is

$$P_0 = \lim_{c\to 0} P_{01} + \lim_{c\to 0} P_{02} = \frac{\alpha_2}{\alpha_1+\alpha_2}\,\frac{\mu_1}{\mu_1+\lambda_1} + \frac{\alpha_1}{\alpha_1+\alpha_2}\,\frac{\mu_2}{\mu_2+\lambda_2}.$$

Thus, the limiting system functions as the (time) average of two independent $M/M/1$ loss systems, one with arrival rate λ_1 and service rate μ_1, and the other with arrival rate λ_2 and service rate μ_2.

4. RIGHT AND WRONG ARRANGEMENTS

Let us assume $\lambda_1 < \lambda_2$ and $\mu_1 < \mu_2$, and compare the system R with levels (λ_1,μ_1), (λ_2,μ_2) to the system W with levels (λ_1,μ_2), (λ_2,μ_1) under the condition $\alpha_1 = \alpha_2$. In other words, the system R has the arrangement such that the server with slow service rate goes on the shift with the slow arrival rate and the person with the fast service rate goes on the shift with the fast arrival rate. The system W is arranged the other way around. If we denote the loss functions of the systems R and W by L_R and L_W, respectively, then a simple algebraic computation yields that $L_R(c) < L_W(c)$, and so the system R is better than the system W in the sense of loss function.

5. FINAL REMARKS

The model considered here is similar to ones considered in Refs. [1] and [3]. However, the results obtained in these papers, being in terms of the root of some polynomial, do not seem to enable one to draw the type of conclusion obtained in the present paper.

REFERENCES

[1] Purdue, P., "The Single Server Queue in a Markovian Environment," in *Mathematical Methods in Queueing Theory*, A. B. Clarke, Editor, (Lecture Notes in Economics and Mathematical Systems, pps. 359-364, Springer-Verlag Publishers, New York 1974).

[2] Ross, S., "Average Delay in Queues with Nonstationary Poisson Arrivals," ORC 77-13, Operations Research Center, University of California, Berkeley (May 1977).

[3] Yechiali, U., and P. Naor, "Queueing Problems with Heterogeneous Arrivals and Service," Operations Research, *19*, 722-734 (1971).

OPTIMAL DISPATCHING STRATEGIES FOR VEHICLES HAVING EXPONENTIALLY DISTRIBUTED TRIP TIMES*

Kamran Asgharzadeh and G. F. Newell

University of Califiornia, Berkeley
Berkeley, California

ABSTRACT

A transportation system has N vehicles with no capacity constraint which take passengers from a depot to various destinations and return to the depot. The trip times are considered to be independent and identically distributed random variables. The dispatch strategy at the depot is to dispatch immediately, or to hold any returning vehicles with the objective of minimizing the average wait per passenger at the depot, if passengers arrive at a uniform rate.

Optimal control strategies and resulting waits are determined in the special case of exponentially distributed trip time for various N up to $N = 15$. For $N >> 1$, the nature of the solution is always to keep a reservoir of vehicles in the depot, and to decrease (increase) the time headway between dispatches as the size of the reservoir gets larger (smaller). For sufficiently large N, one can approximate the number of vehicles in the reservoir by a continuum and obtain analytic expressions for the optimal dispatch rate as a function of the number of vehicles in the reservoir. For the optimal strategy, it is shown that the average number of vehicles in the depot is of order $N^{1/3}$. These limit properties are expected to be quite insensitive to the actual trip time distribution, but the convergence of the exact properties to the continuum approximation as $N \to \infty$ is very slow.

1. INTRODUCTION

We are concerned here with strategies for dispatching vehicles from a depot of a public-transportation system. The system consists of N vehicles, each of which takes passengers from the depot to various destinations, and then returns after some random trip time to make another trip. Passengers arrive at the depot at a constant rate, and vehicles have sufficient capacity that a passenger can always board the next departing vehicle. The objective is to minimize the long-time average wait per passenger at the depot.

A similar problem was previously considered by Osuna and Newell [4], with emphasis on the case of small N, particularly $N = 1$ or 2. For $N = 2$, the coefficient of variation of the trip time, $C(T) = \mathrm{Var}^{\frac{1}{2}}(T)/E(T)$, was further assumed to be small compared to 1. The following analysis will approach the problem from the opposite extremes, cases of $N >> 1$ and/or trip times with large variances, where vehicles are likely to pass each other enroute.

*This research was supported in part by the National Science Foundation under grant MPD 72-05068A03.

It was shown by Osuna and Newell that if the dispatch times $\tau_1 < \tau_2 < \ldots$ define a sequence of headways

$$H_i = \tau_i - \tau_{i-1}, \quad i = 2, 3, \ldots,$$

which are identically distributed random variables satisfying a law of large numbers

$$\frac{1}{n} \sum_{i=1}^{n} H_i \to E(H),$$

and if passengers arrive at a uniform rate independent of dispatch times, the average wait per passenger is

$$E(W) = E(H^2)/2E(H) = (1/2)E(H)[1 + C^2(H)]. \tag{1.1}$$

It is therefore desirable to keep both the mean headway and the variance of H small.

We assume here that the stochastic properties of the trip times are given and that the only mechanism for control is to delay a dispatch. The minimum value of $E(H)$ is achieved by dispatching vehicles with no delay, but then the random trip times generally lead to a large value of $C^2(H)$. The optimal control strategy involves an increase of the "effective trip time" in order to improve the regularity. For small N, this strategy typically requires that one merely delay a vehicle's departure if it returns too early, but there will seldom be more than one vehicle at the depot. For $N >> 1$, however, the optimal strategy will typically involve maintaining a reservoir of vehicles at the depot, so that one will almost always have a vehicle to dispatch at some desired dispatch time. The actual number of vehicles in the reservoir will fluctuate and may occasionally vanish. The optimal strategy describes, among other things, what average fraction of the total number of vehicles one should keep in the reservoir in order to stabilize the departures.

Our analysis of the problem is in two parts. First we give an exact formulation of the problem for exponentially distributed trip times. An analytic solution, although conceptually simple, becomes unmanageable if N becomes large (even $N = 4$), but we do show numerical solutions for $N \lesssim 15$. In the second part we derive an approximate forumlation for the problem, the solution of which gives the asymptotic properties for $N \to \infty$. The numerical and the asymptotic solutions agree fairly well for $N = 15$. For $N >> 1$, the optimal expected number of vehicles kept at the depot is shown to be proportional to $N^{1/3}$.

Although the postulate of an exponentially distributed trip time is not very realistic for most real transportation systems (which typically have a coefficient of variation in trip time considerably less than 1), the results obtained here should give a crude upper bound on the optimal number of vehicles kept at the depot for systems with smaller variation in trip time. For sufficiently large N, however, one can argue that the process of returning vehicles should be approximately a Poisson process, regardless of the detailed nature of the trip time distribution or the dispatch strategy. The optimal control strategy should, therefore, be insensitive to the trip time distribution provided that the uncertainty in trip time is large compared with the mean headway (so that a typical vehicle passes or is passed by many others before it returns). Unfortunately the convergence to such an asymptotic behavior is very slow and of questionable accuracy for any reasonable values of N and $C(T)$ encountered in real systems.

2. Formulation

Let T_i be the trip time of the vehicle which leaves the depot at time τ_i, $i = 1, 2, \ldots$. The T_i's are assumed to be independent identically distributed random variables with a continuous distribution function $F_T(z) = P(T_i \leqslant z)$. We can measure time in any arbitrary units.

Hereafter, all times will be interpreted to be measured in units of $E(T)/N$ (the average headway with no control). In these units

$$E(T)\,N = 1. \tag{2.1}$$

At any time, the information that is relevant to the future behavior of the system is the number of vehicles, x, in the depot; the number of passengers, M, waiting to be served; and the times $t_1 \geqslant t_2 \geqslant \ldots \geqslant t_{N-x}$ since the last dispatches of all the vehicles enroute. Since there is nothing to be gained by dispatching more than one vehicle at a time, the t_i's will satisfy $t_1 > t_2 > \ldots > t_{N-x}$.

We assume that at any time, we know the values of $x, t_1, \ldots, t_{N-x}$, but we cannot observe M. Since passengers arrive at a uniform rate, we will base our strategy on the expected value of M, $E(M)$, which is proportional to the time t since the last dispatch. (The time t since the last dispatch may be equal to one of the t_i or it might be a time since the previous dispatch of one of the x vehicles in the depot.) Since the optimal strategy does not depend upon the arrival rate of passengers, we arbitrarily take this rate to be 1.

We can define a state of the system as the vector $<x, t, t_1, \ldots, t_{N-x}>$. If at all times we follow a strategy (dispatch or not) which depends only upon the present state of the system, the future behavior will depend only upon the present state, and hence we have a Markov process.

Generally, this Markov process is quite complicated. If, however, we assume that the trip times of the vehicles are exponentially distributed, then the future arrival times of the vehicles at the depot are independent of previous dispatch times. The future behavior of the system depends only on the state vector $<x, t>$. This state space, as shown in Figure 1, consists of $N+1$ real lines, $t \geqslant 0$, $x = 0, 1, \ldots, N$.

Since the dispatch strategy depends only on the state of the system, it is specified by a set of points in the state space at which one should dispatch a vehicle, and a complement set of points at which one should hold every vehicle. It is intuitively obvious that if we should dispatch whenever the state is $<x, t>$, then we should also dispatch whenever the state is $<x, t'>$, for $t' \geqslant t$. Therefore, we can assign numbers a_x for each x such that we hold every vehicle if we are at any point $<x, t>$, $t < a_x$, and we dispatch one if we are at a point

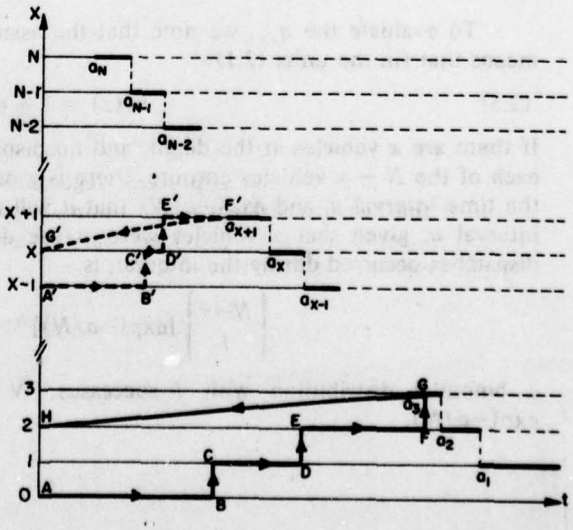


FIGURE 1 — The state space. The light solid lines $(0, a_i)$ are the set of hold points, the dark solid lines $[a_i, a_{i-1}]$ are the set of dispatch points, and the light dashed lines (a_{i-1}, ∞) are the set of unreachable points. Two typical state trajectories ABCEDFG and A' B' C' D' E' F' G' are drawn with dark solid lines and dark dashed lines, respectively.

$<x, t>$, $t \geq a_x$. Note that $a_0 = \infty$ is the only possible value for a_0. It is also obvious that if we should dispatch at a point $<x, t>$, then we should dispatch at a point $<x+1, t>$ because we can better afford to use the vehicles if we have more of them. It follows that

$$\infty = a_0 \geq a_1 \geq a_2 \geq \ldots \geq a_N > 0. \tag{2.2}$$

Figure 1 also shows two types of state trajectories. As long as no departure occurs, the state moves at unit speed from left to right with upward jumps of magnitude one at each arrival of a vehicle (as at BC, DE, B'C', —). At any departure time, however, there is a downward step by one and a return to $t = 0$ (as in GH or F'G'). A dispatch will occur when the state $<x, t>$ first reaches a point with $t \geq a_x$. This may occur either because a vehicle arrived at time t and $a_x < t < a_{x-1}$, as at point G, or the state moved to the point $<x, a_x>$ continuously as at F'. It is not possible to reach states $<x, t>$ with $t > a_{x-1}$ shown by the broken lines; dispatches occur only in the intervals $[a_x, a_{x-1}]$.

The problem now is to find that strategy, i.e., a sequence $a_1^*, \ldots, a_N^*$ among all the possible $a_1 \geq a_2 \geq \ldots \geq a_N$, such as to minimize the expectation of wait of a randomly chosen passenger. Under a strategy of no control, a vehicle is dispatched as soon as it returns no matter how short the headway, i.e., $a_1 = \ldots = a_N = 0$, $a_0 = \infty$. The only states which can be reached, however, are those with $x = 1$ or 0.

For any strategy $\{a_i\}$, the states of the system $<x, 0>$, $x = 0, 1, \ldots, N-1$, immediately after a dispatch define a finite-state Markov process. If $q_{x,y}$ is the conditional probability of having y vehicles after the kth dispatch, given that we had x vehicles after the $(k-1)$th dispatch, and $p_x(k)$ is the probability of having x vehicles after the kth dispatch, then the $p_x(k)$ satisfy the equations

$$p_y(k+1) = \sum_{x=0}^{N-1} q_{x,y}\, p_x(k), \quad y = 0, \ldots, N-1. \tag{2.3}$$

After sufficient time ($k \to \infty$), the $p_y(k)$ will approach an equilibrium distribution p_y which satisfies the equations

$$p_y = \sum_{x=0}^{N-1} q_{x,y}\, p_x, \quad \sum_{x=0}^{N-1} p_x = 1. \tag{2.4}$$

To evaluate the $q_{x,y}$, we note that the assumption of exponentially distributed trip times means that (in the units (2.1))

$$F_T(z) = 1 - \exp(-z/N). \tag{2.5}$$

If there are x vehicles at the depot, and no dispatch occurs during a time interval a, then, for each of the $N - x$ vehicles enroute, there is a probability $1 - \exp(-a/N)$ that it will return in the time interval a, and $\exp(-a/N)$ that it will not. The probability of exactly l arrivals in an interval a, given that x vehicles were at the depot at the beginning of the interval and no dispatches occurred during the interval, is

$$\binom{N-x}{l} [\exp(-a/N)]^{N-x-l} [1 - \exp(-a/N)]^l, \tag{2.6}$$

a binomial distribution with l successes, $N - x - l$ failures, and probability of failure $\exp(-a/N)$.

From Figure 1, one can readily see that a transition from the state $<x, 0>$ to $<\sigma, 0>$, with $0 \leqslant \sigma \leqslant y-1$, will occur if and only if at most $y-x$ vehicles arrive in the time interval $[0, a_y]$. The probability of this event is, therefore,

$$(2.7)\quad Q_{x,y} \equiv \sum_{\sigma=0}^{y-1} q_{x,\sigma} = \sum_{\sigma=0}^{y-x} \binom{N-x}{\sigma} [\exp(-a_y/N)]^{N-x-\sigma} [1-\exp(-a_y/N)]^{\sigma}$$

$$= F_B(y-x;\ N-x,\ 1-\exp(-a_y/N))$$

$$\text{for } x-1 \leqslant y \leqslant N-1,\ x = 0, \ldots, N-1$$

$$= 0 \text{ for } y < x-1,$$

in which $F_B(x;\ n,\ p)$ is the cumulative binomial probability distribution [5], and

$$(2.8)\qquad q_{x,y} = Q_{x,y+1} - Q_{x,y}$$

The solution of the system of N linear equations (2.4), with $q_{x,y}$ given by (2.7) and (2.8), determines the p_x as functions of the strategy $\{a_x\}$. One can also determine the stationary headway distribution

$$(2.9)\qquad \tilde{F}(h) = P(H > h)$$

in terms of the p_x and a_x.

For $h < a_N$, $\tilde{F}(h) = 1$, whereas for $a_{x+1} \leqslant h < a_x$, $x = 0, 1, \ldots, N-1$, it is given by

$$(2.10)\quad \tilde{F}(h) = \sum_{\beta=0}^{x} p_\beta P(H > h \,|\, \beta \text{ vehicles present at the start of } H)$$

$$= \sum_{\beta=0}^{x} p_\beta P(\text{at most } x-\beta \text{ arrivals } in\ (0,\ h) \,|\, \beta)$$

$$= \sum_{\beta=0}^{x} p_\beta F_B(x-\beta;\ N-\beta,\ 1-\exp(-h/N)).$$

From this, one can evaluate

$$(2.11)\qquad E(H) = \int_0^\infty \tilde{F}(h)\, dH,\ E(H^2) = 2\int_0^\infty h\ \tilde{F}(h)\, dh,$$

and

$$E(W) = E(H^2)/2E(H)$$

as functions of the $\{a_x\}$. The problem is thus reduced to minimizing $E(W)$ with respect to the a_x, with $a_1 \geqslant a_2 \geqslant, \ldots \geqslant a_N$.

Except for small N, the evaluation of $E(W)$ and a_x^* is extremely tedious, but one can obtain some bounds on $E(W)$ and the a_x^*. We first note that, if we dispatch with no control, the arrivals of vehicles form a Poisson process with $C(H) = 1$ and, from (1.1), $E(W) = E(H) = E(T)/N = 1$. Therefore

$$E^*(W) \equiv E(W \,|\, a_1^*, \ldots, a_N^*) \leqslant 1.$$

On the other hand, from (1.1), we have

$$E(W) = \frac{E(H)}{2}[1 + C^2(H)] \geqslant \frac{E(H)}{2}.$$

One can show that the minimum value of $E(H)$ is achieved with no control, i.e., $E(H) \geqslant E(T/N) = 1$. Consequently,

$$1/2 \leqslant E^*(W) \leqslant 1. \tag{2.12}$$

The lower bound would obtain if one could maintain regular headways with negligible control delays. The goal for any finite N is to make $E^*(W)$ as close to 1/2 as possible.

Although $E(W)$ is a complex function of $a_1, \ldots, a_{N-1}$, its dependence upon a_N is relatively simple. For any fixed values of $a_1, \ldots, a_{N-1}$, one can show from (2.10), (2.11) that

$$\frac{\partial}{\partial a_N} \frac{E(H^2)}{2} = a_N \frac{\partial E(H)}{\partial a_N},$$

and that the equation

$$\partial E(W)/\partial a_N = 0$$

is satisfied provided that a_N satisfies the equation

$$a_N = E(W). \tag{2.13}$$

Thus, for any choice of $a_1, \ldots, a_{N-1}$, the optimal a_N must satisfy (2.13). This does not give a_N "explicitly", because $E(W)$ depends upon a_N, but $E(W)$ varies so slowly with a_N that a sequence of successive approximations will converge very rapidly.

Equation (2.13) can also be derived by a dynamic-programming argument, such as described in reference [4], for the special case $N = 1$.

Since we know that $E^*(W) \geqslant \frac{1}{2}$, it follows that

$$\frac{1}{2} \leqslant E^*(W) = a_N^* \leqslant a_{N-1}^* \leqslant \ldots \leqslant a_0^* = \infty. \tag{2.14}$$

Also, the strategy $a_N = a_{N-1} = \ldots = a_1 = \frac{1}{2}$ will give a smaller $E(W)$ than no control, $a_N = a_{N-1} = \ldots = a_1 = 0$, therefore a better upper bound for $E^*(W)$ than (2.12).

3. NUMERICAL RESULTS FOR SMALL N

For $N = 1$ and 2 (and perhaps 3), evaluation of the optimal control can be done analytically, but for larger N the solution of (2.4) and the minimization of $E(W)$ with respect to the a_x must be done numerically.

The solution for $N = 1$ was described in reference [4] even for a general trip-time distribution. The optimal strategy is to dispatch the single vehicle immediately if it arrives after some time a_1, but dispatch at time a_1 if it arrives before time a_1. Equation (2.4) has the trivial solution $p_0 = 1$. The headway distribution (2.10) is a truncated exponential

$$\tilde{F}(h) = \begin{cases} 1, & 0 < h \leqslant a_1, \\ \exp(-h), & a_1 < h, \end{cases}$$

giving

$$E(H) = a_1 + \exp(-a_1), \text{ and } E(H^2) = a_1^2 + 2(a_1 + 1)\exp(-a_1).$$

The optimal a_1 is given by (2.13)

$$a_1^* = \frac{a_1^{*2}/2 + (a_1^* + 1)\exp(-a_1^*)}{a_1^* + \exp(-a_1^*)}, \tag{3.1}$$

which can be solved by successive approximations. If we substitute any jth approximation $a_1^*(j)$ in the right-hand side of (3.1) and evaluate the $(j+1)$th approximation as the value of the left-hand side, this sequence of approximations will converge very rapidly. For example, if we take a trial value of $a_1^*(0) = 1/2$, then $a_1^*(1) = 0.935$ and $a_1^*(2) = 0.901$ (already correct to three decimal places).

As compared with no control, the optimal control has reduced $E(W)$ from 1 to approximately 0.90. This reduction is achieved through increasing $E(H)$ from 1 to 1.307 and decreasing $C^2(H)$ from 1 to 0.38. This illustrates how the average wait can be reduced through an improvement in the regularity in service at the expense of longer average headways.

For $N = 2$, there are two control parameters a_1 and a_2. The transition probabilities (2.7) become

$$Q_{0.1} = q_{0.0} = 1 - q_{0.1} = 2\exp(-a_1/2) - \exp(-a_1)$$

and

$$Q_{1.1} = q_{1.0} = 1 - q_{1.1} = \exp(-a_1/2),$$

and the solution of (2.4) is

$$p_0 = 1 - p_1 = \frac{\exp(-a_1/2)}{1 - \exp(-a_1/2) + \exp(-a_1)},$$

which is independent of a_2.

The headway distribution (2.10) becomes

$$\tilde{F}(h) = \begin{cases} 1, & 0 < h < a_2, \\ (1 - p_0)\exp(-h/2) - p_0\exp(-h), & a_2 < h < a_1 \\ p_0\exp(-h), & a_2 < h \end{cases}$$

from which one can evaluate $E(H)$, $E(H^2)$, and $E(W)$ as explicit functions of a_1 and a_2.

The pair of equations

$$\frac{\partial E(W)}{\partial a_1} = \frac{\partial E(W)}{\partial a_2} = 0 \text{ for } a_1 = a_1^* \text{ and } a_2 = a_2^*,$$

for the optimal a_1 and a_2, can be manipulated into the form

$$a_1^* = -2\ln[\exp(-a_2^*/2) - a_2^{*2}/8], \quad a_2^* = E^*(W). \tag{3.2a,b}$$

The minimum of $E(W)$ can again be obtained by successive approximations. If we substitute any jth approximation $a_2^*(j)$ in (3.2a), we obtain a jth approximation $a_1^*(j)$. The $(j+1)$th approximation for a_2^* is then obtained by substituting $a_1^*(j)$ and $a_2^*(j)$ into $E(W)$, as in (3.2b). This sequence of approximations converges very rapidly. Table 1 illustrates the iterations starting from $a_2^*(0) = 1/2$.

TABLE 1

	Iteration Number		
	0	1	2
a_2^*	0.500	0.899	0.838
a_1^*	0.582	1.243	1.125
p_0^*	0.922	0.715	0.755
$E^*(H)$	1.091	1.366	1.310
$E^*(W)$	0.899	0.838	0.835

Whereas the optimal control for $N = 1$ reduced $E(W)$ from 1 to 0.901, $E(W)$ can be further reduced to 0.835 for $N = 2$. Again, as in the case of one vehicle, this result is obtained by an increase of $E(H)$ from 1 to 1.31 (approximately the same as for $N = 1$) and by a larger decrease in $C^2(H)$, from 1 to 0.28.

For $N > 2$, it becomes progressively more tedious to determine the a_j^*s by setting derivatives of $E(W)$ equal to zero. One should, instead, use numerical techniques better suited for a computer. For any N and any particular choice of the a_j's, the matrix $q_{x,y}$ was computed numerically from (2.7) and (2.8). The system of N linear equations (2.4) was then solved for the p_x. That $q_{x,y} = 0$ for $y < x - 1$ means that the yth equation of (2.4) can be solved for p_{y+1} in terms of $p_0, \ldots, p_y$, $y \leqslant N - 1$. Thus, for $y = 0, 1, \ldots$, one can successively evaluate $p_1, p_2, \ldots$, in terms of p_0, then determine p_0 from the normalization. The headway distribution can be evaluated from (2.10), and the values of $E(H)$, $E(H_2)$, and $E(W)$ calculated.

The evaluation of $E(W)$ for any particular N and $\{a_x\}$ is computationally quite simple even for moderately large N. To minimize $E(W)$ with respect to the $\{a_x\}$, however, we exploited certain known properties of the a_x^*, such as (2.14), and followed an iterative scheme for which $E(W)$ would converge rapidly to $E^*(W)$. The scheme successively minimizes $E(W)$ with respect to one variable at a time in the following way.

First we take $a_2(0) = a_3(0) = \ldots = a_N(0) = 0.5$ and let a_1 vary over the interval [0.5, 2.5] (in steps of 0.1). We find the best $E(W)$, $E(W)$ (0), and the corresponding $a_1(0)$. Next we choose $a_1(1) = a_1(0)$ and $a_3(1) = a_4(1) = \ldots = a_N(1) = E(W)(0)$ and let a_2 vary over the interval $[a_3(1), a_1(1)]$. We find the best $E(W)$, $E(W)(1)$, and the corresponding $a_2(1)$. In the third step, we choose $a_1(2) = a_1(1)$ $a_2(2) = a_2(1)$, and $a_4(2) = \ldots = a_w(2) = E(W)(1)$ and minimize $E(W)$ with respect to a_3.

We continue until we have minimized $E(W)$ with respect to a_{N-1}, chosen $a_N = E(W)$ $(N - 2)$, and obtained a new trial sequence $a_N(N - 1) \leqslant a_{N-1}$ $(N - 2) \leqslant \ldots \leqslant a_1(0)$. Starting with this trial solution, we now repeat the above procedure of successively determining new values of $a_1, a_2, \ldots, a_N$. As was true in the numerical schemes for $N = 1$ and 2, this procedure converges within only a few cycles.

It took only 1.6 minutes for a CDC 6400 computer to run the program for $N = 15$. This time would likely increase rapidly with N. Our purpose here, however, was mostly to bridge the

gap between the results for $N = 1$ and 2 and the asymptotic results described in the next sections. No attempt was made to develop more efficient programs for dealing with larger N because the practical limitations of the model do not justify high precision.

Results of these calculations are shown in Figures 2, 3, 4, and 5. Figure 2 shows values of $E^*(W)$ as a function of N. Although these are defined only for integer N, the points at integer N have been joined by a smooth curve. The unit of time (2.1) has been chosen as the average uncontrolled headway, not the trip time. If one were simply to add more vehicles to an existing route, keeping $E(T)$ constant, then the average wait in real time units would be $E^*(W)E(T)/N$ which will, of course, decrease with N at a much faster rate that $E^*(W)$. The reason why $E^*(W)$ decreases with N is that, the larger N is, the less the control of any one headway affects the future arrivals. All uncontrolled systems give a Poisson arrival process of vehicles, but the larger N is, the larger is the space of control strategies. For $N \to \infty$, the optimal control will permit nearly regular headways, and $E_B^*(W) = 1/2$, as compared with the uncontrolled strategy with $E_N(W) = 1$.

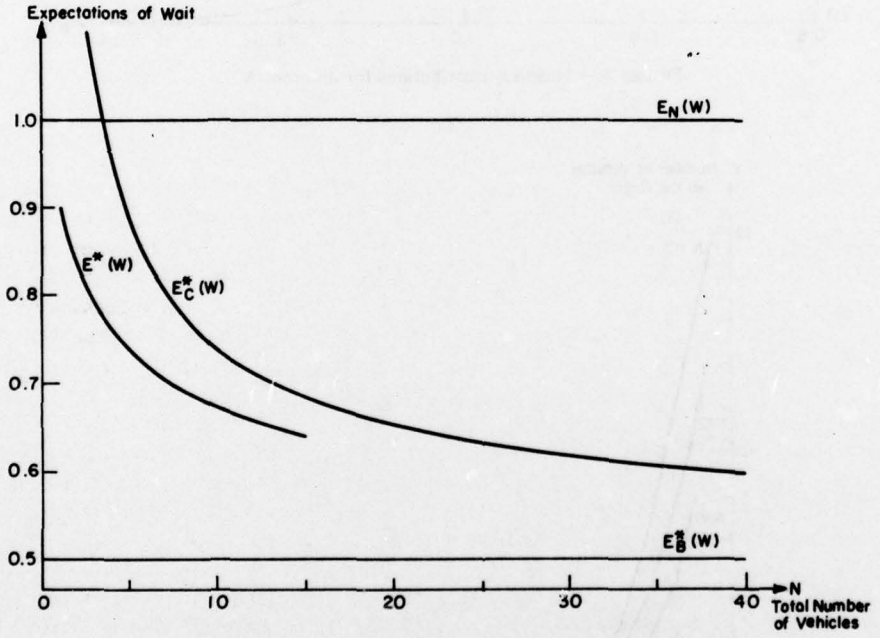


FIGURE 2 — Different expectations of wait versus N.

Figure 3 illustrates the distribution of controlled headways for $N = 1$, 2, 3, 15, and ∞, and also the common exponential distribution for the case of no control. We expect that, for $N \to \infty$, the optimal control will produce $H = 1$ with probability 1 because, by keeping only a finite number of vehicles at the depot, one can make the probability of running out of vehicles (p_0) arbitrarily small. One can see, however, that the approach of $\tilde{F}^*(h)$ toward that of $N = \infty$ is quite slow. Figure 3 also shows an approximate $\tilde{F}(h)$ for $N = 1000$ obtained from asymptotic formulas derived in the next sections.

Figure 4 illustrates the optimal strategies a_x^* for several values of N. Since vehicles are dispatch at a rate $1/E(H)$, this is also the rate of return. For small x, $x = 1, 2, \ldots$, one

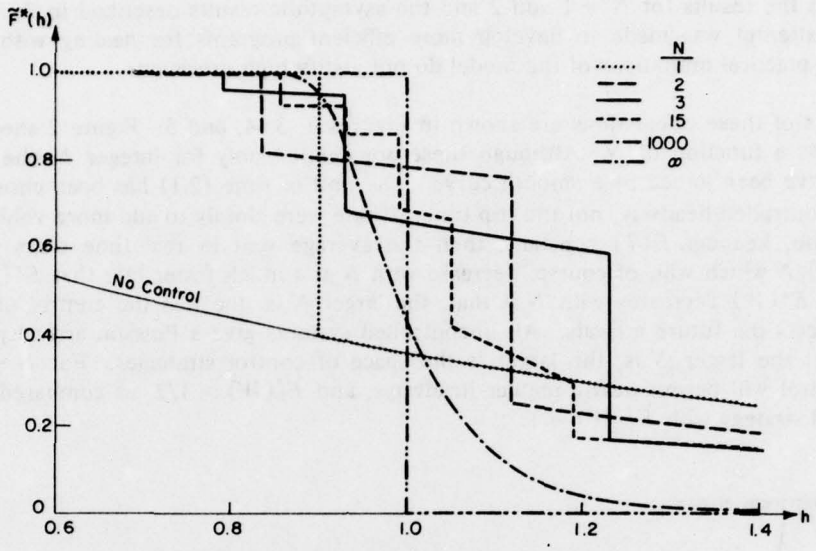


FIGURE 3 — Headway distributions for different N.

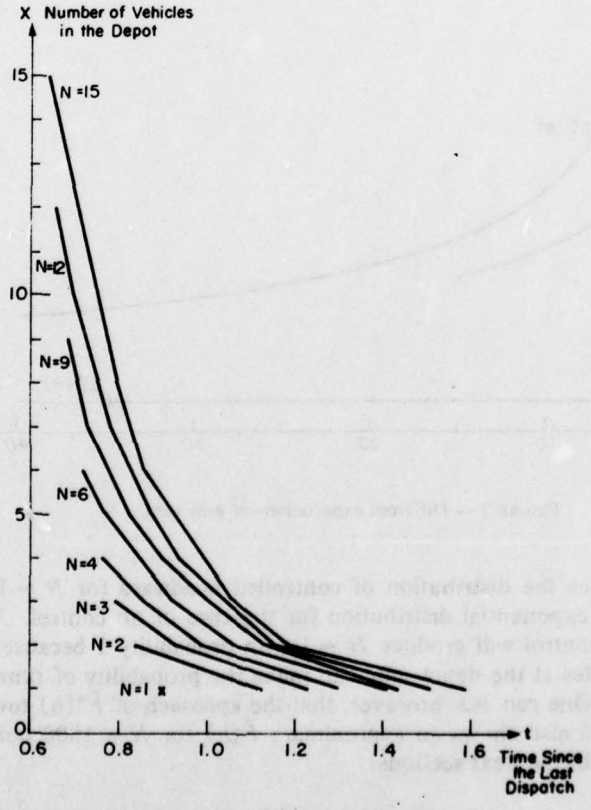


FIGURE 4 — Control strategies for different N.

dispatches with headways larger than $E(H)$ to prevent the state from reaching 0, but if one has many vehicles in the depot, one dispatches them more frequently.

Figure 5 shows the distribution of the number of vehicles available in the depot immediately after a dispatch for the cases of $N = 1, 2, 3$, and 15. The distribution is drawn with coordinate $y = N^{-1/3}x$ in anticipation of the limit behavior for $N \to \infty$ described in the next sections.

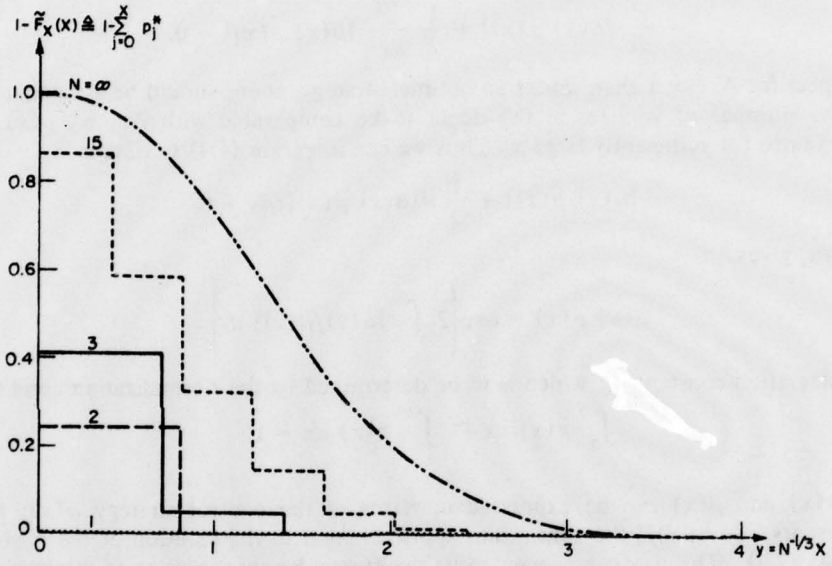


FIGURE 5 — Distribution of the number of vehicles in the depot immediately after a dispatch.

4. A CONTINUOUS APPROXIMATION

Figures 3, 4, and 5 suggest that for $N >> 1$, the a_x^*'s and p_x^*'s change relatively little as x changes by 1. It seems advantageous, therefore, to change the domain of x from the integers $\{0, 1, \ldots, N\}$ to the continuous interval $[0, N]$ and to assume that a_x of Figures 1 or 4 can be approximated by a smooth curve $a(x)$ and that $p_x(k)$ can be approximated by a smooth function $p(x, k)$. The $p(x, k)$, which can be interpreted as a "probability density" of the number of vehicles at the depot immediately after the kth dispatch, will satisfy a continuous counterpart or (2.3)

$$p(y, k+1) = \int_0^N q(x,y)\, p(x,k)\, dx, \tag{4.1}$$

in which $q(x,y)$ is the conditional probability density of having y vehicles after the $(k+1)$th dispatch, given that there were x vehicles after the kth dispatch.

We anticipate that, for $N >> 1$, the effective "width" of the distribution $p(x,k)$ will be large compared with the probable one step changes of the state, $y - x$, and that the $p(x, k)$ will, therefore, approximately satisfy a diffusion equation of the form [2]*

$$p(x, k+1) - p(x, k) = -\frac{\partial}{\partial x}[\alpha(x)p(x,k)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\beta(x)p(x, k)], \tag{4.2}$$

*In reference [2], the left-hand side of (4.2) would be written as $\partial p(x,k)/\partial k$.

in which

$$\alpha(x) \equiv E(Y - x|x) \text{ and } \beta(x) \equiv E((Y - x)^2|x) \tag{4.3}$$

are the first and second moments about x of the random variable $Y|x$ having a probability density $g(x, y)$.

For $k \to \infty$, we expect $p(x, k)$ to approach an equilibrium distribution $p(x)$ which is a solution of the ordinary differential equation

$$-\frac{d}{dx}[\alpha(x)\ p(x)] + \frac{1}{2}\frac{d^2}{dx^2}[\beta(x)\ p(x)] = 0. \tag{4.4}$$

We also expect for $N >> 1$ that, under an optimal strategy, there should be a negligible probability for the number of vehicles at the depot to be comparable with N, i.e., $p(x)$ and any derivatives vanish for sufficiently large x. Thus we can integrate (4.4) to obtain

$$-[\alpha(x)\ p(x)] + \frac{1}{2}\, d[\beta(x)\ p(x)]/dx = 0, \tag{4.5}$$

which in turn, gives

$$\beta(x)\ p(x) = \exp\left\{2\int_{x_0}^{x} [\alpha(y)/\beta(y)]\ dy\right\} \tag{4.5a}$$

for some integration constant x_0, which is to be determined by the normalization condition

$$\int_0^N p(x)\ dx \approx \int_0^\infty p(x)\ dx = 1. \tag{4.5b}$$

The $\alpha(x)$ and $\beta(x)$ can be expressed in terms of the control strategy $a(x)$; therefore (4.5a) represents, essentially, the continuum approximation to the solution of the N simultaneous equations (2.4). The $a(x)$, however, must eventually be chosen so as to minimize $E(W)$, which can also be expressed in terms of the $p(x)$ and $a(x)$.

Actually, instead of using (4.5a) to express $p(x)$ in terms of $a(x)$, so as to minimize $E(W)$ with respect to $a(x)$, we will use (4.5) to express $a(x)$ in terms of $p(x)$ and minimize $E(W)$ with respect to $p(x)$. Whereas in the discrete case we minimized $E(W)$ with respect to the N parameters $a_1, a_2, \ldots, a_N$, the continuum version will involve minimizing $E(W)$ with respect to a function $a(x)$ or $p(x)$, and give a calculus of variation problem.

5. CALCULATION OF $\alpha(x)$ AND $\beta(x)$

The exact transition probabilities $q_{x,y}$ are given in terms of a_y by (2.7) and (2.8). Since a_x is to be approximated by a smooth function, we expect that $a_x - a_{x+1}$ will be small compared with a_x. It is convenient therefore to note that transition from x to $x + l$ can occur from any of the mutually exclusive events

A $\equiv$ exactly l+1 arrivals in $[0, a_{x+l}]$

B $\equiv$ exactly l+1 arrivals in $[0, a_{x+l+1}]$ and 1 or more arrivals in $[a_{x+l+1}, a_{x+l}]$, or

C $\equiv$ at most l arrivals in $[0, a_{x+l+1}]$ and at least l+2 arrivals in $[0, a_{x+l}]$.

Event B involves at least one arrival in $[a_{x+l+1}, a_{x+l}]$, and C involves at least two. For small values of $a_{x+l} - a_{x+l+1}$, we can write

$$p(A) = \binom{N-x}{l+1} [1 - \exp(-a_{x+l}/N)]^{l+1} [\exp(-a_{x+l}/N)]^{N-x-l-1},$$

$$p(B) = \binom{N-x}{l+1} [1 - \exp(-a_{x+l+1}/N)^{l+1} [\exp(-a_{x+l+1}/N)]^{N-x-l-1}$$

$$\times (a_{x+l} - a_{x+l+1}) + 0(a_{x+l} - a_{x+l+1})^2$$

$$= (a_{x+l} - a_{x+l+1})\, p(A) + 0(a_{x+l} - a_{x+l+1})^2,$$

and

$$p(C) = 0(a_{x+l} - a_{x+l+1})^2,$$

therefore

$$q_{x,x+l} = \binom{N-x}{l+1} [1 - \exp(-a_{x+l}/N)]^{l+1} [\exp(-a_{x+l}/N)]^{N-x-l-1} \cdot [1 + (a_{x+l} - a_{x+l+1})] + 0(a_{x+l} - a_{x+l+1})^2, \quad l \geqslant -1. \tag{5.1}$$

The conditional moments $\alpha(x)$ and $\beta(x)$ can be evaluated from

$$\alpha(x) = \sum_{l=-1}^{N-x-1} l q_{x,x+l} \text{ and } \beta(x) = \sum_{l=-1}^{N-x-1} l^2 q_{x,x+l}, \tag{5.2}$$

but these are difficult to evaluate exactly, because a_{x+l} is a function of l. To simplify them, we expand $a_{x+l} = a(x+l)$ as

$$a_{x+l} = a(x+l) = a(x+1) + (l-1)a'(x+l) + \frac{(l-1)^2}{2} a''(x+1) + \ldots,$$

where

$$a'(x+1) \equiv \frac{d}{dx} a(x+1) \text{ and } a''(x+1) = \frac{d^2}{dx_2} a(x+1).$$

If we expand the terms of $\alpha(x)$ and $\beta(x)$ in powers of $a'(x+1)$, $a''(x+1)$, x, x/N, a_x/N, etc., the sum over l can be evaluated term by term from the moments of a Poisson distribution. For purposes of combining terms of the same order of magnitude, however, it is convenient here to anticipate the relative magnitudes of x, $a'(x+l)$, etc. We assume now, and verify later, that the relevant range of x is $x = 0(N^{1/3})$ for $N >> 1$, and that in this range of x (except possibly for $x = o(N^{1/3})$),

$$a(x+1) = 1 + 0(N^{-1/3}),\ a'(x+1) = 0(N^{-2/3}),\ a''(x+1) = 0(N^{-1}). \tag{5.3}$$

That $a(x+1)$ should be approximately one derives from the fact that one must have an average of one arrival between each dispatch to maintain an equilibrium. An $a(x) > 1$ means that vehicles are, on the average, returning faster than they are being dispatched.

An expansion of (5.2) in powers of N gives, for $N >> 1$,

$$\alpha(x) = a(x+1) - 1 - (x/N) + 0(N^{-1}) \tag{5.4}$$

and

$$\beta(x) = a(x+1) + 0(N^{-2/3}) = 1 + 0(N^{-1/3}). \tag{5.5}$$

6. THE EXPECTATION OF WAIT

To express $E(H)$, $E(H^2)$, and $E(W)$ in terms of $a(x)$, $p(x)$, we write $E(H)$ in the form

$$E(H) = \sum_{j=0}^{N} \sum_{i=0}^{j+1} E(H|i,\ j)\ q_{i,j}\ p_i, \tag{6.1}$$

where $E(H|i,j)$ is the expectation of a headway that starts with i vehicles, and ends with j vehicles after the dispatch, and $q_{i,j}p_i$ is the joint probability of having i and j such vehicles.

Since the headway that ends with j vehicles after the dispatch is usually a_{j+1}, and a_{j+1} is close to one for most j, we rewrite (6.1) in the form

$$E(H) = \sum_{j=0}^{N} \sum_{i=0}^{j+1} [1 + (a_{j+1} - 1) + E(H - a_{j+1}|i,j)]\ q_{i,j}\ p_i.$$

We know that

$$\sum_{i=0}^{j+1} q_{i,j}\ p_i = p_j, \text{ and } E(H - a_{j+1}|j+1,j) = 0,$$

therefore

$$E(H) = 1 + \sum_{j=0}^{N} (a_{j+1} - 1)p_j + \sum_{j=0}^{N} \sum_{i=0}^{j} E(H - a_{j+1}|i,j)q_{i,j}p_i. \tag{6.2}$$

Similarly,

$$E(H^2) = 1 + \sum_{j=0}^{N} (a_{j+1}^2 - 1)p_j + \sum_{j=0}^{N} \sum_{i=0}^{j} E(H^2 - a_{j+1}^2|i,j)\ q_{i,j}\ p_i. \tag{6.3}$$

We anticipate that the first term of (6.2) will be large compared with subsequent terms. If we expand $1/E(H)$ about one, and multiply it by $E(H^2)$, we obtain

$$\begin{aligned} 2E(W) &= 1 + \sum_{j=0}^{N} (a_{j+1}^2 - 1)p_j - \sum_{j=0}^{N} (a_{j+1} - 1)p_j \\ &\quad + \sum_{j=0}^{N} \sum_{i=0}^{j} [E(H^2 - a_{j+1}^2|i,j) - E(H - a_{j+1}|i,j)]q_{i,j}\ p_i + R \\ &= 1 + \sum_{j=0}^{N} (a_{j+1} - 1)a_{j+1}\ p_j \\ &\quad + \sum_{j=0}^{N} \sum_{i=0}^{j} E[(H - a_{j+1})\ (H + a_{j+1} - 1|i,j]\ q_{i,j}\ p_i + R \end{aligned} \tag{6.4}$$

in which R contains terms proportional to second or higher powers of $[E(H) - 1]$ and $[E(H^2) - 1]$.

In order for H to deviate from a_{j+1}, it is necessary that there be at least one arrival in $[a_{j+1},\ a_j]$. Therefore,

$$\begin{aligned} &E[(H-a_{j+1})\ (H+a_{j+1} - 1)|i,j] \\ &= E[(H-a_{j+1})\ (H+a_{j+1} - 1)|i,j \text{ and } H \neq a_{j+1}]\ P\ (H \neq a_{j+1}|i,j) \end{aligned} \tag{6.5}$$

is of order $(a_j - a_{j+1})^2$. Except possibly for the first few values of j, say $j < \gamma = o(N^{1/3})$, we expect from (5.3) that $(a_j - a_{j+1})^2 \sim [a'(j)]^2 = 0(N^{-4/3})$. These terms will give a negligible contribution to (6.4).

The terms in (6.5) for small i,j are not necessarily negligible, particularly the term for $i = j = 0$, because $a_0 = \infty$. If we started with no vehicles in the depot and no vehicle returned by a_1, we would have to wait until the arrival of the first vehicle. If this extra waiting occurs, we would wait on the average another average arrival headway. This term of (6.5) represents the penalty for running out of vehicles, which a proper control strategy will want to avoid.

We could consider separately the terms of (6.4) that contain the p_j with $j > \gamma$ from those with $j > \gamma$, and write (6.4) in the form

$$2E(W) = 1 + \sum_{j=\gamma}^{N} (a_{j+1} - 1)\, a_{j+1} p_j + \sum_{j=0}^{\gamma-1} b_j\, p_j + R + 0(N^{4/3}), \tag{6.6}$$

in which the b_j are positive functions of the a_j's.

For $j > \gamma$, we can approximate the a_{j+1} and p_j by the continuous functions $a(x+1)$ and $p(x)$ and write (6.6) as

$$2E(W) \approx 1 + \int_{\gamma-1/2}^{\infty} [a(x+1) - 1]a(x+1)\, p(x)\, dx + \sum_{j=0}^{\gamma-1} b_j\, p_j + R. \tag{6.7}$$

The second term of (6.7) involves both the $a(x+1)$ and $p(x)$, but these two functions are related through (4.5), (5.4), and (5.5). We can use these equations to express $a(x+1)$ in terms of $p(x)$. From (5.4) and (4.5), we have

$$a(x+1) - 1 \approx \alpha(x) + \frac{x}{N} = \frac{1}{2p(x)} \frac{d}{dx} [\beta(x)\, p(x)] + \frac{x}{N} + 0(N^{-1}).$$

Thus the integrand of (6.7) becomes

$$\begin{aligned} p(x)\, a(x+1)\, [a(x+1) - 1] &\approx p(x)\, [a(x+1) - 1] + p(x)\, [a(x+1) - 1]^2 \\ &\approx \frac{1}{2} \frac{d}{dx} [\beta(x)p(x)] + \frac{x}{N} p(x) + p(x) \left\{ \frac{1}{2p(x)} \frac{d}{dx} [\beta(x)\, p(x)] + \frac{x}{N} \right\}^2. \end{aligned} \tag{6.8}$$

The first term of this is of the largest order relative to N, but it is a perfect differential and integrates to

$$\begin{aligned} \int_{\gamma-1/2}^{\infty} \frac{1}{2} \frac{d}{dx} [\beta(x)\, p(x)]\, dx &= -\frac{1}{2} p\left(\gamma - \frac{1}{2}\right) \left[1 + \alpha\left(\gamma + \frac{1}{2}\right) - 1\right] \\ &\approx -\frac{1}{2} p(\gamma - \tfrac{1}{2}) \left[1 + \frac{1}{2\, p(\gamma - 1/2)} \frac{d}{d\gamma} p(\gamma - \tfrac{1}{2})\right] \approx -\frac{1}{2} p(\gamma). \end{aligned}$$

In the last term of (6.8), we can approximate $\beta(x)$ by one and neglect the x/N, so that (6.7) becomes

$$2E(W) \approx 1 + \int_{\gamma-1/2}^{\infty} \left[\frac{x}{N} p(x) + \frac{1}{4p(x)} \left[\frac{dp(x)}{dx}\right]^2\right] dx - \frac{1}{2} p(\gamma) + \sum_{j=0}^{\gamma-1} b_j p_j + R. \tag{6.9}$$

Although we have used the equations relating the a_j's to the p_j's to eliminate $a(x+1)$ from the integral of (6.7), the p_j, b_j, and a_j for $j < \gamma$ are still constrained to satisfy (2.4). The details of this are quite complex, but the issues are clear.

The p_j must satisfy the normalization condition

$$\sum_{j=0}^{\gamma-1} p_j + \int_{\gamma-1/2}^{\infty} p(x)\, dx \approx 1, \tag{6.10}$$

and, through (2.4), the $p(\gamma)$ in (6.9) depends upon the p_j, $0 \leqslant j \leqslant \gamma-1$. By appropriate control, we can distribute the probabilities p_j so that either most of the probability lies in $0 < j < \gamma - 1$, or most lies in $j > \gamma$. The term $-(1/2)p(\gamma)$ in (6.9) suggests that, to minimize (6.9), one should assign most of the probability to small j so as to make this term as large as possible. One can show that this term is identified with the fact that the smaller j is, the more vehicles there are in use, and consequently the shorter the average headway is.

On the other hand, the terms $b_j p_j$ of (6.9) for $0 \leqslant j \leqslant \gamma-1$ are all positive, particularly the term for $j = 0$, which includes an extra wait if $j = 0$ and no vehicle has returned by the time a_1, when a dispatch is desired. To minimize these terms, one should have a very small probability for $j \leqslant \gamma - 1$. One can show that the effect of these terms overpowers the term $-(1/2)p(\gamma)$, and dictates a policy for which the p_j for $j < \gamma - 1$ are small for $N >> 1$.

For sufficiently larger N, the strategy which minimizes $E(W)$ must be approximately one for which $p(x)$ is chosen so as to minimize the second term of (6.9) with $\gamma \approx 0$, i.e., minimize

$$J = \int_0^{\infty} \left[\frac{x}{N}\, p(x) + \frac{1}{4p(x)} \left[\frac{dp(x)}{dx} \right]^2 \right] dx, \tag{6.11}$$

subject to

$$p(0) = 0 \text{ and } \int_0^{\infty} p(x)\, dx = 1. \tag{6.11a,b}$$

7. LIMIT DISTRIBUTION

To minimize (6.11) subject to (6.11a,b) is a standard calculus of variation problem. It can be transformed into a more familiar form, however, if we let

$$p(x) = \frac{1}{N^{1/3}} f^2\left[\frac{x}{N^{1/3}} \right], \quad y = \frac{x}{N^{1/3}}. \tag{7.1}$$

Then, in terms of $f(y)$, minimize (6.11) subject to (6.11a,b) becomes the well known Sturm-Liouville Problem [3]:
minimize

$$J = \frac{1}{N^{2/3}} \int_0^{\infty} \left[y\, f^2(y) + \left[\frac{df(y)}{dy} \right]^2 \right] dy \tag{7.2}$$

subject to

$$f(0) = 0 \text{ and } \int_0^{\infty} f^2(y)\, dy = 1. \tag{7.2a,b}$$

The Euler equation corresponding to minimize (7.2) subject to (7.2b) is

$$\frac{d^2 f^*(y)}{dy^2} = (y + \lambda)\, f^*(y). \tag{7.3}$$

which is the Sturm-Liouville differential equation with Lagrange multiplier λ. The only solution of (7.3) that vanishes for $y \to \infty$ is

$$f^*(y) = K\,\mathrm{Ai}\,(y + \lambda)$$

in which K is a constant and $\mathrm{Ai}(\cdot)$ is the Airy Function [1]. The normalization (7.2b) determines K; thus

$$f^*(y) = \mathrm{Ai}(y + \lambda)/\{\int_\lambda^\infty \mathrm{Ai}^2(\xi)\, d\xi\}^{1/2} \tag{7.4}$$

and

$$p^*(x) = \frac{1}{N^{1/3}}\,\mathrm{Ai}^2\left(\frac{x}{N^{1/3}} + \lambda\right)\Big/\int_\lambda^\infty \mathrm{Ai}^2(\xi)\, d\xi. \tag{7.5}$$

This determines $f^*(y)$, except for a translation of coordinates by λ, but the boundary condition (7.2a) requires that $\mathrm{Ai}(\lambda) = 0$. Figure 6 shows graphs of $\mathrm{Ai}(z)$, $\mathrm{Ai}^2(z)$ and

$$\mathrm{Ai}'(z)/A(z) = d \ln \mathrm{Ai}(z)/dz.$$

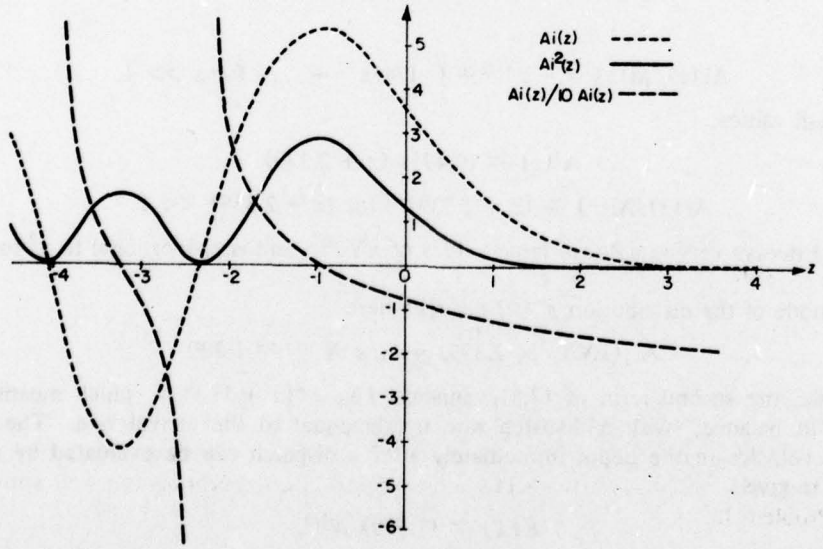


FIGURE 6 — Airy functions.

Although $\mathrm{Ai}(z)$ has zeros at $z \approx -2.339, -4.088, \ldots$, we know from (4.5) and (5.4) that

$$a(N^{1/3}y + 1) \approx 1 + \frac{1}{2p(x)}\,\frac{dp(x)}{dx} = 1 + \frac{d}{dy} \ln \mathrm{Ai}(y + \lambda),$$

and that $a(x)$ must be a monotone decreasing function of x. As one can see from Figure 6, values of $y + \lambda$ less than -2.339 would violate the latter condition, consequently λ must be the zero at

$$\lambda \approx -2.339. \tag{7.6}$$

The normalization in (7.4) and (7.5) can be evaluated by numerical integration:

$$\int_{-2.339}^{\infty} \mathrm{Ai}^2(\xi)\, d\xi \approx 0.491.$$

Thus the optimal limit distribution and control are

$$p^*(x) \approx 2.037\, N^{-1/3}\, \mathrm{Ai}^2(xN^{-1/3} - 2.339) \tag{7.7}$$

and

$$a^*(x+1) \approx 1 + N^{-1/3}\, \mathrm{Ai}'(xN^{-1/3} - 2.339)/\mathrm{Ai}(xN^{-1/3} - 2.339). \tag{7.8}$$

This verifies the conjecture made in (5.3), particularly that $a(x+1) = 1 + 0(N^{-1/3})$. Indeed the dependence of $p^*(x)$ and $a^*(x+1) - 1$ upon N involves merely a scaling of coordinates with all lengths x measured relative to $N^{1/3}$. For large values of $xN^{-1/3}$, one can use the asymptotic approximations

$$\mathrm{Ai}(z) = \frac{1}{2\pi^{1/2}}\, z^{1/4} \exp\left[-\frac{2}{3}\, z^{3/2}\right]\left[1 - \frac{5}{48}\, z^{-3/2} + \ldots\right] \tag{7.8a}$$

and

$$\mathrm{Ai}'(z)/\mathrm{Ai}(z) \approx -z^{+1/2} + (-1/4)z^{-1} + \ldots, \text{ for } z \gg 1, \tag{7.8b}$$

and, for small values,

$$\mathrm{Ai}(z) \approx (0.491)\,(z + 2.339) \tag{7.9a}$$

$$\mathrm{Ai}'(z)/\mathrm{Ai}(z) \approx (z + 2.339)^{-1} \text{ for } (z + 2.339) \ll 1. \tag{7.9b}$$

Thus, $p^*(x)$ decays very rapidly for large values of $xN^{-1/3}$, and is proportional to x^2 for small x.

The mode of the distribution $p^*(x)$ occurs where

$$\mathrm{Ai}'(xN^{-1/3} + 2.399) = 0,\ x\, N^{-1/3} \approx 1.320.$$

At the mode, the second term of (7.8) vanishes, i.e., $a^*(x+1) \approx 1$, which means that the system is "in balance," with a dispatch rate nearly equal to the arrival rate. The expected number of vehicles in the depot immediately after a dispatch can be evaluated by numerical integration to give

$$E(X) \approx (1.559)\, N^{1/3},$$

somewhat larger than the mode because of the skewed distribution.

Of the N vehicles one has available, the fraction of vehicles which one keeps at the depot is thus of order $N^{-2/3}$.

The smooth curve of Figure 5 shows the complementary distribution function

$$\int_x^{\infty} p^*(x')\, dx' = \int_{xN^{-1/3}}^{\infty} \mathrm{Ai}^2(y' - 2.339)\, dy'$$

as a function of $xN^{-1/3}$. One can see that the discrete distributions for $N = 1$, 3, and 15 are approaching this limit distribution but, since these asymptotic expansions are essentially expansions in powers of $N^{-1/3}$, one does not expect very rapid convergence relative to N. The main error seems to be associated with the fact that, for finite N, one can allow a nonzero probability for $x = 0$ with a moderately large a_1. For $N = 15$, p_0 is still about 0.14. To obtain a second approximation in the asymptotic expansions is very tedious; attempts to do so were not very fruitful. One can see from Figure 5 that a translation $x \to x + 1$ would improve the fit considerably, but the asymptotic expansions cannot be expected to distinguish between $x = 0$ and $x = 1$.

Figure 7 compares a_x^* for $N = 15$ with the continuous $a^*(x)$. On this scale, $0 \leqslant x \leqslant 15$, the agreement does not look very good, but most of the state probability lies in the range $x = 1$ to 5, where the agreement is satisfactory.

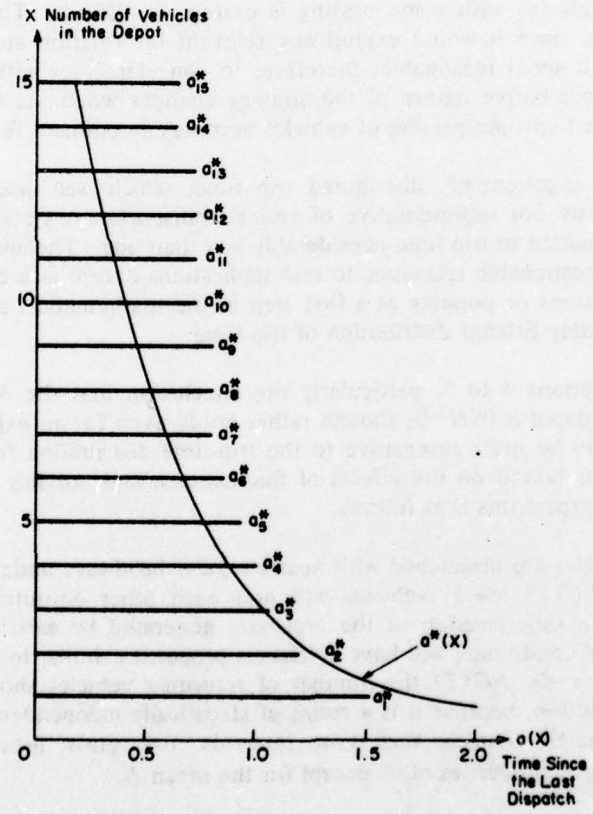


FIGURE 7 — The optimal strategy for $N = 15$.

Since the headway which ends with x vehicles in the depot is approximately $a^*(x + 1)$, and the $p^*(x)$ are known, one can easily evaluate a continuum approximaton to the headway distribution $F^*(h)$. The curve of Figure 3 for $N = 1000$ is obtained in this way. This clearly demonstrates how slowly this distribution converges to the limit $H \equiv 1$.

By substituting (7.4) into (7.2), one can evaluate the integral $JN^{2/3}$, and show that

$$2E^*(W) \approx 1 + (2.282)\ N^{-2/3} + o(N^{-2/3}).$$

In Figure 2, this continuum approximation is represented by the curve labeled $E_C^*(W)$.

8. DISCUSSION

The main purpose of the above analysis was to investigate the nature of efficient control strategies for a public-transportation system having a large number N of vehicles serving the same depot. The measure of performance is considered to be the average wait of passengers at the depot.

Previous work [4] has described strategies (at least for small N) in which the trip time of vehicles is so predictable that vehicles do not pass enroute. The analysis of optimal control strategies for several vehicles with some passing is extremely difficult. The optimal strategy would be very complex since it would exploit any relevant information about past departure times of all vehicles. It seem reasonable, therefore, to consider cases with large N as some indication of how the qualitative nature of the strategy changes when the fluctuations in trip time become so large that enroute passing of vehicles becomes a dominant feature.

The postulate of exponentially distributed trip time, which was made to simplify the mathematics, is obviously not representative of typical transportation systems, which usually have a coefficient of variation in trip time considerably less than one. The analysis of sections 2 and 3 is therefore of questionable relevance to real applications except as a crude upper bound on the effect of fluctuations or possibly as a first step in the mathematical analysis of systems with more general (possibly Erlang) distribution of trip time.

The results of Sections 4 to 7, particularly the conclusion that the optimal number of vehicles to keep at the depot is $0(N^{1/3})$, though rather crude even for an exponential trip-time distribution, are likely to be quite insensitive to the trip-time distribution for $N >> 1$. They should provide an upper bound on the effects of fluctuations, and possibly a close one. The reason that one should expect this is as follows.

Even though vehicles are dispatched with nearly regular headways under the optimal control, if $N >> 1$ and $NC(T) >> 1$, vehicles will pass each other enroute. The process of returning vehicles is the superposition of the processes generated by each of the N vehicles and, under quite general conditions, will have stochastic properties similar to a Poisson process. In particular, in a time $\tau << NC(T)$ the number of returning vehicles should have approximately a Poisson distribution, because it is a count of statistically independent rare events (the return of any ith vehicle). During such time intervals, the return process is, therefore, insensitive to the stochastic properties of T, except for the mean N.

From (4.2) one can show that the "natural unit of time" for the diffusion equation is of order $N^{2/3}$. This is the time it takes for the system to reach an equilibrium or to return to an equilibrium after some disturbance. If the return process behaves like Poisson process for times of this order, i.e., if $\tau \approx N^{2/3} << NC(T)$ or $C(T) >> N^{-1/3}$, one might expect the diffusion equation to be approximately correct. Of course, $N^{-1/3}$ is not very small for reasonable values of N.

The results described here may be useful as a bound on the strategies for real systems, but they are not expected to be quantitatively accurate for any of the systems which motivated this analysis. The intended applications were for bus routes (downtown terminal to suburbs, or

airport to downtown) with N perhaps about 20 and $C(T)$ probably about 0.1, or elevators with N about 6 and a $C(T)$ perhaps about 0.3. Unfortunately, these values of N are not large enough for the present theory to be accurate.

REFERENCES

[1] Abramowitz, M., and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

[2] Cox, D. R., and H. D. Miller, *The Theory of Stochastic Processes* (John Wiley, New York, 1965).

[3] Gelfand, I. M., and S. V. Fomin, *Calculus of Variations*, (Prentice Hall, Englewood Cliffs, N.J. 1963).

[4] Osuna, E. E., and G. F. Newell, "Control Strategies for an Idealized Public Transportation System," *Transportation Science*, *6*, 52-72 (1972).

[5] *Tables of the Binomial Probability Distribution*, National Bureau of Standards Applied Mathematics Series 6, (U.S. Government Printing Office, Washington, D.C., 1950).

EVALUATION OF COMMONLY USED RULES FOR DETECTING "STEADY STATE" IN COMPUTER SIMULATION*

A. V. Gafarian, C.J. Ancker, Jr., and T. Morisaku

Department of Industrial and Systems Engineering
University of Southern California
Los Angeles, California

ABSTRACT

A definition of the problem of the initial transient with respect to the steady-state mean value has been formulated. A set of criteria has been set forth by which the efficacy of any proposed rule may be assessed. Within this framework, five heuristic rules for predicting the approximate end of transiency, four of which have been quoted extensively in the simulation literature, have been evaluated in the M/M/1 situation. All performed poorly and are not suitable for their intended use.

1. INTRODUCTION

In this paper we consider some of the problems posed by the existence of an initial transient in the system response which may arise in a digital-computer simulation of a stochastic process. Basically, the situation is this. Suppose that a discrete-parameter stochastic process $\{X_t,\ t = 1, 2, 3, \ldots\}$ is being observed for which a set of initial conditions, denoted by I, exist at $t = 0$. For example, X_t may be inherently discrete, like the waiting time of the t^{th} customer arriving at a queuing system after the simulation begins; or, it may arise by sampling, at equidistant time intervals, a continuous time series such as the number of jobs in a system. We suppose that the first moments of these random variables exist and tend to an asymptotic limit, independent of I, i.e.,

$$\lim_{t \to \infty} E[X_t | I] = \mu_\infty,$$

where μ_∞ is defined as the steady-state mean. We assert that the principal problem of the initial transient is that of determining the minimum t, call it t^*, such that the expectation of the random variables X_t, $t \geqslant t^*$, is as close as one desires to the limiting expectation.† Symbolically, t^* is the smallest t for which

$$1 - \epsilon \leqslant \frac{E[X_t]}{E[X_\infty]} \leqslant 1 + \epsilon, \qquad t \geqslant t^*,$$

*This research was supported in part by the National Science Foundation under Grant Number ENG 75-06900.

†Throughout this paper we use the standard notation of an upper-case letter for a random variable and a lower-case letter for a specific value the random variable may assume.

where $E[X_\infty] = \mu_\infty$ is the steady-state expected value and $\epsilon > 0$ is a preassigned number. Thus, for example, $\epsilon = 0.05$ if one desires to be within five percent of the steady-state mean value. Figure 1 illustrates examples of some ways that $E[X_t]/E[X_\infty]$ may converge and also the t^* corresponding to the given ϵ. Though our notation does not show it explicitly, it should be noted that t^* depends on ϵ.

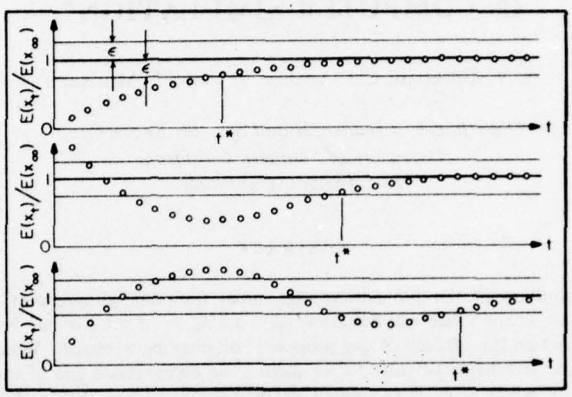


FIGURE 1 — Some ways $E[X_t]/E[X_\infty]$ converges and the associated t^* value for the indicated ϵ value.

The importance of this problem arises from the fact that a simulator must know the conditions under which data are being collected. Is the process near the limiting condition of steady state, or does it still have a long way to go to be near the equilibrium condition? That question must be answered in order to collect and analyze data appropriate for any specific purpose. For example, if all data are collected far from equilibrium they cannot be used to produce good estimates of the steady-state mean. The disastrous consequences of not properly accounting for the initial transient have been illustrated very well by Law [5] in the construction of a confidence interval for the steady-state mean, using either the method of replication or the method of batch means. Knowledge of t^* is not only necessary for estimation of steay-state parameters, but also for parameters of the transient part of the process. In the latter case, repetitive samples are required, and it is wasteful to go past t^* or inadequate to be far short of it.

Several methods or rules of thumb for determining an estimate $\hat{t}^*$ of t^* have been suggested in the literature and are discussed later. It was anticipated that these methods for detecting t^* would state a very specific procedure for obtaining a unique estimate $\hat{t}^*$ of t^* in any particular instance. However, a careful examination of existing rules, whose origins are always based on some rationale, shows that such a goal is difficult to attain in every case. First, the procedure itself is not always unambiguously defined; sometimes, certain parameters are left to the investigator to select. Second, even when these parameters are selected, the application of the rule does not necessarily result in a unique estimate — i.e., certain judgmental aspects remain, so that two people looking at the same set of data and carrying out identical procedures will come up with different estimates for t^*. To remove these vagaries and attain unambiguous specificity would require substantial effort to pin these rules down any further. As a matter of fact, the performance characteristics of all of them are so poor that to make them more explicit would be a useless endeavor.

In the remainder of this paper we (1) discuss the criteria for goodness of a detection rule, (2) describe some rules that exist in the literature and a slight modification of one, and, finally, (3) present the results of an empirical evaluation study of these rules.

2. GOODNESS CRITERIA

In assessing the goodness of a rule for detecting t^* or estimating t^*, there are several desirable characteristics to look for. These are accuracy, precision, generality, low cost, and simplicity. Each of these will be discussed in turn; but first the reader should be reminded again that, given a rule (including its vagaries), we are merely estimating a well-defined number t^*. Thus, many of the concepts associated with estimation theory, such as unbiasedness and mean square error, apply.

2.1. Accuracy

A rule is a statement that tells us how to obtain a $\hat{t}^*$. Of course, this is just one possible value of a random variable $\hat{T}^*$, i.e., $\hat{T}^*$ is an estimator of t^*. Accuracy will be used as a measure of location. Thus, it seems that an appropriate definition of accuracy, a, would be

$$a = \frac{E[\hat{T}^*]}{t^*}.$$

If this ratio is close to one, then we say the detector is accurate; if it is greater than one it implies a positive bias, and less than one, a negative bias.

2.2 Precision

Precision, p, will be used as a measure of variation; more specifically, the coefficient of variation of $\hat{T}^*$, namely,

$$p = \frac{\sqrt{\text{Var}[\hat{T}^*]}}{E[\hat{T}^*]}$$

will be defined as the precision of the estimator. Clearly, a small value is desirable; when p is close to zero we say that the detector is precise.

2.3. Generality

This is a property which means that the rule performs well across a broad range of systems and a broad range of parameters within a system.

2.4. Cost

By cost we mean the expense, in computer time, from using a given detection rule. As will be seen, there are three factors which we combine to arrive at a total cost. Not all factors appear in every rule. These factors are

(i) computer time for the algorithm itself, i.e., its computational efficiency,

(ii) computer time for collecting computer output data only for a preliminary estimate of t^*, and subsequently discarding this data, and

(iii) computer time associated with a positively biased rule. Thus, if $E[\hat{T}^*] >> t^*$, then on any replication of the simulation experiment, one of two situations may occur, unbeknownst to the analyst. Either (a) more data is generated than would be required if one were studying the transient situation, or (b) for estimating a steady state parameter, μ_∞ for instance, the useful data generated between t^* and the smallest integer greater than or equal to $E[\hat{T}^*]$ would not be used.

2.5. Simplicity

This is a characteristic of a rule which makes it accessible to the average practitioner of large-scale system simulation. A rule utilizing abstruse mathematical or statistical results is nearly incomprehensible, and virtually useless, to the average person who needs to know how to get statistically reliable results from a simulation.

In evaluating any specific rule, its accuracy, precision, and generality should be considered first. A rule that is not satisfactory on all three counts is obviously undesirable, and not worth pursuing. Unless a rule is prohibitively expensive, cost should be a relatively minor consideration. However, it is a matter of personal judgement where one balances budget constraints with the necessity for a good predictor. The criterion of simplicity is always satisfied in the rules we are considering in this study.

3. DETECTION RULES

In this section we describe the rules most commonly appearing in the simulation literature. The results of an empirical evaluation study of these rules, in terms of the criteria discussed above, are presented later.

3.1 Rule 1 (Conway Rule)

Conway [1] says, "I usually truncate a series of measurements until the first of the series is neither the maximum nor the minimum of the remaining set. I do not do this for every run, but rather decide on a stabilization period by examining a few pilot runs and thereafter delete this same period from the result of each run."

Thus, this rule specifies a priori the number of exploratory replications and the number of observations per exploratory run, denoted by e and ℓ, respectively. Then if $\{x_{i1}, x_{i2}, \ldots, x_{i\ell}\}$ is the set of observations on the i^{th} exploratory run, one computes

$$x_{ik}^{+} = \max\{x_{ik}, x_{i,k+1}, \ldots, x_{i\ell}\}$$

and

$$x_{ik}^{-} = \min\{x_{ik}, x_{i,k+1}, \ldots, x_{i\ell}\}$$

for $k = 1, 2, \ldots, \ell$ and determines t_i such that

$$x_{it_i}^{-} < x_{it_i} < x_{it_i}^{+}$$

occurs for the first time. Then the estimate of t^* is given by

$$\hat{t}^* = \max\{t_1, t_2, \ldots, t_e\}.$$

A schematic diagram of this situation is shown in Figure 2 for a single exploratory run of length 10.

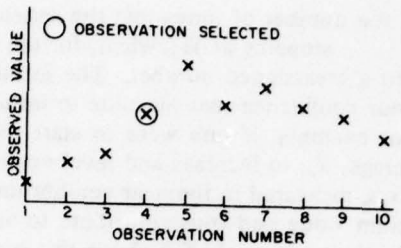


FIGURE 2 — Conway Rule applied to a single exploratory run of length 10.

3.2 Rule 2 (Modified Conway Rule)

This rule is the only one considered in this paper that is not in the literature. In this rule, we turn Conway's idea around and continually look backwards to find the first observation that is neither a maximum or minimum of all the *previous* observations. Thus, the total number of observations in this procedure is a random variable, in contrast to the Conway rule. Again, the method requires a prespecified number of exploratory replications, e, each of which produces a stopping point. The maximum stopping point in this set of stopping points is selected as $\hat{t}^*$.

A schematic diagram of this situation is shown in Figure 3 for a single exploratory run. The observation values shown are the same as those in Figure 2. Note, however, that the simulation stops, in this case with observation 6, when the criterion is met.

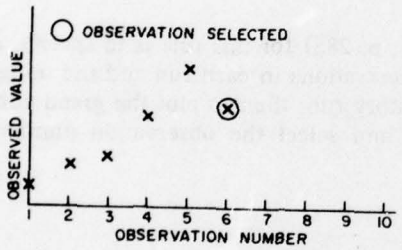


FIGURE 3 — Modified Conway rule applied to a single exploratory run.

3.3. Rule 3 (Crossings-of-the-Mean Rule)

This rule, appearing in Fishman (Ref. [2], p. 275), specifies that a running cumulative mean be computed as the data are generated, and that a count be made of the number of crossings of the mean, looking backwards to the beginning. When the number of crossings reaches a prespecified number, then one has arrived at $\hat{t}^*$. Thus, if the segment $\{x_1, x_2, \ldots, x_n\}$ has been generated, define

$$\omega_j = \begin{cases} 1, \text{ if } x_j > \bar{x}_n,\ x_{j+1} < \bar{x}_n \text{ or } x_j < \bar{x}_n,\ x_{j+1} > \bar{x}_n, \\ 0, \text{ otherwise} \end{cases}$$

with $j = 1, 2, \ldots, n-1$, where

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^{n} x_j.$$

Then compute

$$\Omega_n = \sum_{j=1}^{n-1} \omega_j,$$

which gives us the number of times that the series $x_1, x_2, \ldots, x_n$ crosses the mean. One computes $\Omega_2, \Omega_3, \ldots$, stopping at $\Omega_{\hat{n}}$ when, for the first time, the number of crossings is greater than or equal to a preassigned number. The intuitive notion is that the larger this number is, the greater is our confidence that bias due to initial conditions has been resolved. This seems reasonable. For example, if one were to start the system out empty, one would expect the cumulative average, $\bar{x}_n$, to increase and level out to its equilibrium value. When this happens, the individual x_j's, measured in the near equilibrium situation, would be sprinkled on either side of the equilibrium value and thus contribute to an increasing Ω_n. However, during the early stages, the individual x_j's would fall above this increasing cumulative mean, and therefore add nothing to the Ω_n. A schematic of the situation is shown in Figure 4.

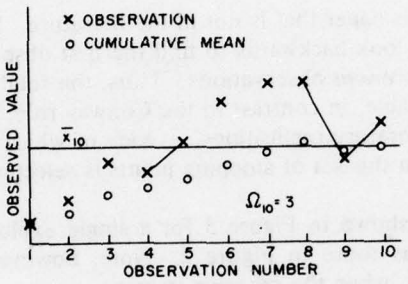


FIGURE 4 — Crossings of the mean.

3.4 Rule 4 (Cumulative-Mean Rule)

The procedure described by Gordon (Ref. [3], p. 285) for this rule is to specify, a priori, the number of exploratory runs, the number of observations in each run and the initial condition for each that is held constant for each exploratory run, then to plot the grand cumulative mean over all exploratory runs and observations and select the observation number which *appears* to be stable.

3.5 Rule 5 (Gordon Rule) (Ref. [3], p. 285)

For a very large class of processes, namely ones for which covariance stationarity obtains and where

$$m = \sum_{j=-\infty}^{\infty} R_j$$

is finite, and where $R_j = R_{-j}$ = covariance of lag j, it can be shown (Ref. [3] p. 281) that

$$\mathrm{Var}[\bar{X}_n] = \frac{m}{n} + o(\frac{1}{n}),$$

where $o(\frac{1}{n})$ denotes terms of order higher than 1/n. However, suppose there exists an initial bias. Label the random variables

$$X_1', X_2', \ldots, X_{n^*}', X_{n^*+1}', \ldots, X_{n^*+n}',$$

where n^* corresponds to the point where near equilibrium is obtained and n is the number of additional observations beyond that point. Define

$$\bar{X}_{n^*+n}' = \frac{1}{n^*+n} \sum_{i=1}^{n^*+n} X_i',$$

$$\bar{X}_{n^*}' = \frac{1}{n^*} \sum_{i=1}^{n^*} X_i',$$

and

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j,$$

where

$$X_j = X_{n^*+j} \qquad j = 1, 2, \ldots, n.$$

Then, it can be shown that

$$\text{Var}[\bar{X}_{n^*+n}] = \left(\frac{n}{n^*+n}\right)^2 \frac{m}{n} + o\left(\frac{1}{n}\right).$$

Thus, after a sufficiently large number of observations beyond n^*, the variance of $\bar{X}_{n^*+n}$ behaves like the variance of $\bar{X}_n$ which has no initial bias. The similarity of behavior is in the sense that the variance consists of a term of order $1/n$ plus a term of order higher than $1/n$.

The idea, proposed by Gordon requires a priori specification of the number of exploratory runs, e, the number of observations to be made in each exploratory run, ℓ, and the initial condition which is held constant for each exploratory run, x_O. Denoting the mean value of the j^{th} exploratory run by $\bar{x}_{j,n}$, one may estimate Var $[\bar{X}_n]$ by

$$s^2(n) = \frac{1}{e-1} \sum_{j=1}^{e} (\bar{x}_{j,n} - \bar{x}_N)^2,$$

where e is the number of exploratory runs,

$$N = ne,$$

$$\bar{x}_N = \frac{1}{e} \sum_{j=1}^{e} \bar{x}_{j,n} = \frac{1}{ne} \sum_{i=1}^{n} \sum_{j=1}^{e} x_{ij},$$

and

$x_{ij} = i^{\text{th}}$ observatioin of the j^{th} exploratory run.

As indicated above, in the absence of initial bias, $s^2(n)$ can be expected to be inversely proportional to n or $s(n)$ inversely proportional to $\sqrt{n}$ for larger values of n; however, in the presence of initial bias, such as our constant initial condition, this behavior will manifest itseft only after the effects of initial conditions are eliminated. Gordon suggests plotting $s(n)$ versus n on log-log paper and finding the point on the plot where it steadies into a straight line with a negative slope of 1/2.

3.6 Comment

A very important observation here is that Rule 3 wastes no data. Thus, in practice, the run for determining $\hat{t}^*$ would be included as part of the simulation, so that in carrying out the experiment, when a point in time is declared as near equilibrium, the experimenter may then stop (if, say, he is conducting a study of transient characteristics), or he may continue from $\hat{t}^*$ on, to collect data while the system is near steady state. The point is that there is no separation of experiments; i.e., one set to determine a near-equilibrium point and a second set of data-collecting experiments. This is not the case for rules 1, 2, 4, and 5, as they are not an integral part of the simulation model. Rather, these rules require that an initial set of experiments be conducted to specifications laid down by the rule, the sole purpose being to provide data for estimating an appropriate stopping point. The actual data-collecting experiments require addi-

tional replications of the simulation. All the data generated for establishing the near-equilibrium point are essentially wasted.

4. EMPIRICAL EVALUATION STUDY

As stated earlier, the criterion of simplicity has been met by all five rules described above. Each of them would be comprehensible to any simulator with a modest background in mathematics and mathematical statistics, and is readily usable. So this point will no longer be addressed.

The next three criteria to be considered are those of accuracy, precision, and generality. The obvious way to go about doing this is to look at how the detection rule performs over a wide spectrum of systems. We began with one of the best-known queuing systems, namely, $M/M/1$. If a rule performs well in this instance, i.e., it is accurate and precise over a wide range of initial conditions x_0 = number of customers in the system at $t = 0$ and parameter values $\rho = \frac{\lambda}{\mu}$, where λ is equal to the arrival rate and μ the service rate, respectively, then it would be necessary to assess the generality of the rule. These considerations, i.e., accuracy and precision, would have to be studied with other systems. This study is only a beginning. The $M/M/1$ queuing system was chosen for the following reasons:

(1) Queuing simulations occur often.

(2) Theoretical values of the quantities of interest can be computed with relative ease.

(3) It is easy and inexpensive to simulate.

In order to carry out our evaluations, it is necessary to determine the true t^* for various x_0 and ρ combinations. This means that the expectation $E[X_t]$ must be known as a function of t as well as $E[X_\infty]$. These are basically analytical problems; and, as pointed out in the preceding paragraph, one reason for selecting $M/M/1$ is that the analysis is tractable.

The observed random variable (chosen from many possible outputs), referred to in general as X_t so far, is, for the M/M/1 model,

W_t = waiting time in queue of the t^{th} arrival before being served.

The simulation was carried out recursively using

$$W_t = \max(0,\ W_{t-1} + S_{t-1} - A_t),$$

where

S_{t-1} = service time of the $(t-1)^{\text{th}}$ arrival,

A_t = interarrival time of the t^{th} arrival.

$E[W_t]$ for an initial condition of empty is given in Ref. [4]. The expectations for nonempty conditions were developed in the course of this study; see Ref. [6] for details.

The main objective of the empircal study was to estimate the accuracy and precision of the various rules. For rules 1, 2, and 3, we made 100 independent estimates $\hat{t}_i^*$, $i = 1,2,\ldots,$ 100, for each value of ρ and x_0 considered and used

$$\hat{a} = \frac{\frac{1}{100}\sum_{i=1}^{100} \hat{t}_i^*}{t^*} = \hat{E}[\hat{T}^*]$$

as the estimate of the accuracy a, and

$$\hat{p} = \frac{(\widehat{\mathrm{Var}}[\hat{T}^*])^{1/2}}{\hat{E}[\hat{T}^*]}$$

as the estimate of precision, p, where

$$\widehat{\mathrm{Var}}[\hat{T}^*] = \frac{1}{100}\sum_{i=1}^{100} (\hat{t}_i^* - \hat{E}[\hat{T}^*])^2,$$

and

$$\hat{E}[\hat{T}^*] = \frac{1}{100}\sum_{i=1}^{100} \hat{t}_i^*.$$

Each of the 100 estimates was determined by running the simulation and selecting $\hat{t}^*$ in accordance with the prescription provided by the rule. This procedure, of replicating 100 times, was too expensive for rules 4 and 5. However, we are confident, based on only one replication, that our conclusions regarding these rules are correct.

4.1 Rule 1 (Conway Rule)

We have followed the spirit of Conway's thought that one should use a "few" exploratory replications and a small number of observations. The specified number of exploratory replications, e, was always either 1, 3, 5, or 10. The specified length of each replication, ℓ (i.e., the number of customers), was varied as follows: 4(1) 10(2) 20(5) 30(10) 50(25) 100. A wide range of parameters ρ and x_0 was tested, as shown in Table 1. In general, it was possible to make 100 replications for each ρ, x_0, e, ℓ combination. However, in some situations, especially for low ℓ and ρ values, a replication may produce a sequence of waiting times for which the Conway criterion is not met , i.e., there exists no observation such that it is neither a minimum or a maximum of all succeeding observations. For example, in the case of $\rho = 0.1$, $x_0 = 0$, $e = 1$, and $\ell = 5$, only one out of the 100 replications was successful in producing a $\hat{t}^*$. Table 2 shows the number of successful replications out of 100, as a function of the exploratory run length ℓ, for the conditions specified previously. Obviously, the estimates of accuracy and precision are based only on the subset of successful runs.

TABLE 1 — *M/M/1 Queuing Parameters Used to Test the Conway Rule*

ρ \ x_0	0	5	10	25
0.1	x	x		
0.5	x	x		
0.7	x		x	
0.9	x			x

TABLE 2 — *Number of Successful Replications out of 100 Using the Conway Rule with $\rho = 0.1$, $x_0 = 0$, and $e = 1$*

	Number of Successful Replications
5	1
10	7
15	18
20	27
25	43
30	55
40	72
50	82
75	97
100	98

The Conway rule resulted in a very poor set of performance characteristics. Figures 5a, b, c, and d show plots of accuracy for $\rho = 0.1$, 0.5, 0.7, and 0.9, respectively, and various values of the other parameters. These plots are all for $\epsilon = 0.10$; however, the results are essentially identical for $\epsilon = 0.05$. They show that for low ρ values this rule overestimates t^*, and for high ρ values it grossly underestimates t^*. Thus, it fails on the accuracy criterion.

Note that $e = 1$ does not appear on Figure 5a. This was because there were not enough successful replications to get good estimates. This is also the reason why the left endpoint is not the same for all curves. The plots begin only when ℓ is sufficiently large to produce enough replications for a good estimate.

Similar plots for estimates of precision were also obtained. Figures 6a and b, for $\rho = 0.5$, are typical of these results — roughly that $0.3 \leqslant p \leqslant 0.6$ for $x_0 = 0$, and $0.4 \leqslant p \leqslant 0.8$ for the nonempty starting case, i.e., when $x_0 = 5$, 10, and 25, for $\rho = 0.1$, 0.5, 0.7, and 0.9, respectively.

4.2 Rule 2 (Modified Conway Rule)

In the original form of the Conway Rule, the length of the exploratory runs, l, must be specified in advance, and one looks ahead only in order to make a determination of whether or not the criterion is met. In the Modified Conway Rule, since one only looks back for making the determination, there is no need to specify an exploratory run length; i.e., the exploratory run continues until the criterion is met and an estimate $\hat{t}^*$ is acquired. Thus, the only parameter in this version of the rule is the number of exploratory runs, e. This also makes presentation of the data substantially simpler.

The performance in this case was very poor also. Figures 7a, b, c, and d summarize the accuracy results for $\epsilon = 0.10$; the results for $\epsilon = 0.05$ are essentially the same. In these particular sets of runs, an additional value of x_0 was run for each value of p. This intermediate value was in some cases selected to be very close to the steady-state number in the system, which is given by $\rho/(1 - \rho)$. Thus, for example, in Figure 7d, the additional value is $x_0 = 10$ and the steady-state value is 9. The Modified Conway Rule badly underestimates t^* in virtually all cases, except for the intermediate x_0 values at $\rho = 0.5$ and $\rho = 0.7$.

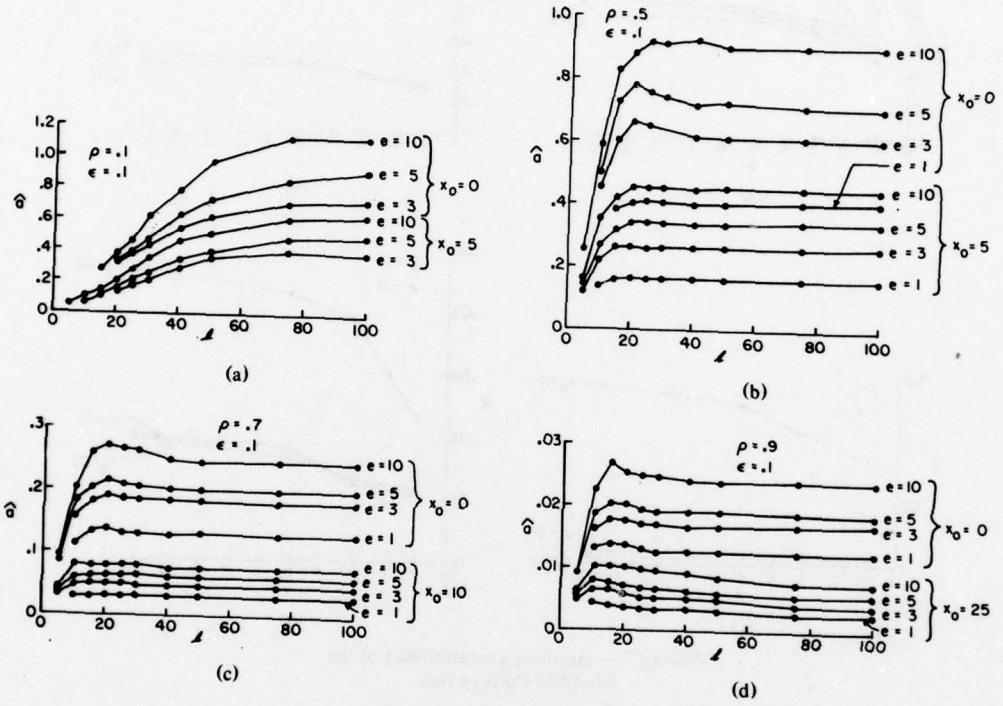


FIGURE 5 — Accuracy characteristics of the Conway Rule.

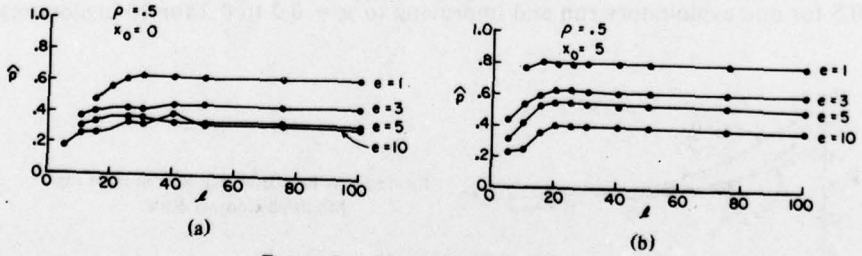


FIGURE 6 — Precision characteristics of the Conway Rule.

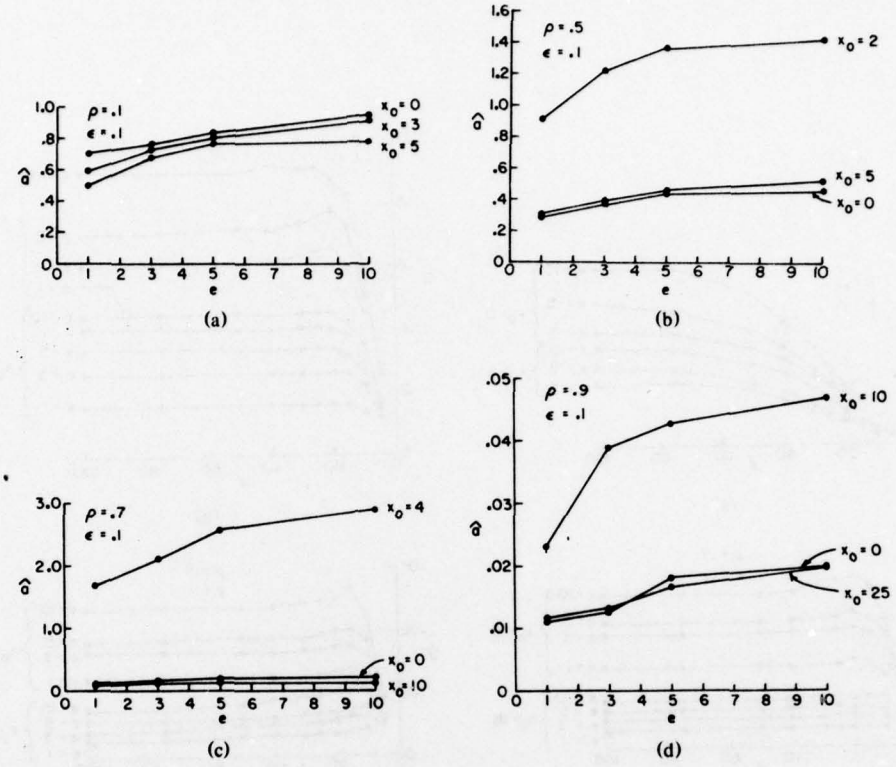


FIGURE 7 — Accuracy characteristics of the Modified Conway rule.

Again, similar plots were obtained for precision characteristics. An example of one is shown in Figure 8 for $p = 0.5$. Its appearance is typical of all curves, starting with about $p = 0.4$ to 0.5 for one explolratory run and improving to $p = 0.2$ to 0.3 for 10 exploratory runs.

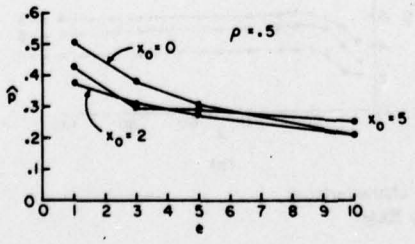


FIGURE 8 — Precision characteristics of the Modified Conway Rule.

As in the Conway Rule, the data obtained in these experiments are probably not used in any subsequent inferential studies and, therefore, may prove to be quite costly.

In summary, both the Conway and Modified Conway Rules have poor performance characteristics with respect to accuracy in the $M/M/1$ situation. These rules will no longer be tested in other situations, since, even if they produced good results in some other cases, our criterion for generality would not have been met.

4.3 Rule 3 (Crossings-of-the-Mean Rule)

In testing this rule, the only open parameter we considered is the number of crossings to be attained before declaring that near-equilibrium had been reached. The range of system parameters was the same as that used in the Conway Rule (see Table 1). Figures 9a, b, c, and d display accuracy versus the number of crossings for $\rho = 0.1, 0.5, 0.7$, and 0.9, respectively. It is clear from these figures that the rule is very conservative for low ρ values, i.e., it overestimates the t^* and provides accuracies much greater than one, no matter what criterion is selected for the number of crossings. As ρ increases it becomes less conservative; in fact, for $\rho = 0.9$ its accuracy is less than one for crossings below 25. Nevertheless, it appears that a crossings criterion of 25 would work uniformly across all system parameters in the sense that it provides an estimate $\hat{t}^*$ which, on the average, is greater than t^* for all conditions tested. Thus, the system is closer than expected to the equilibrium condition.

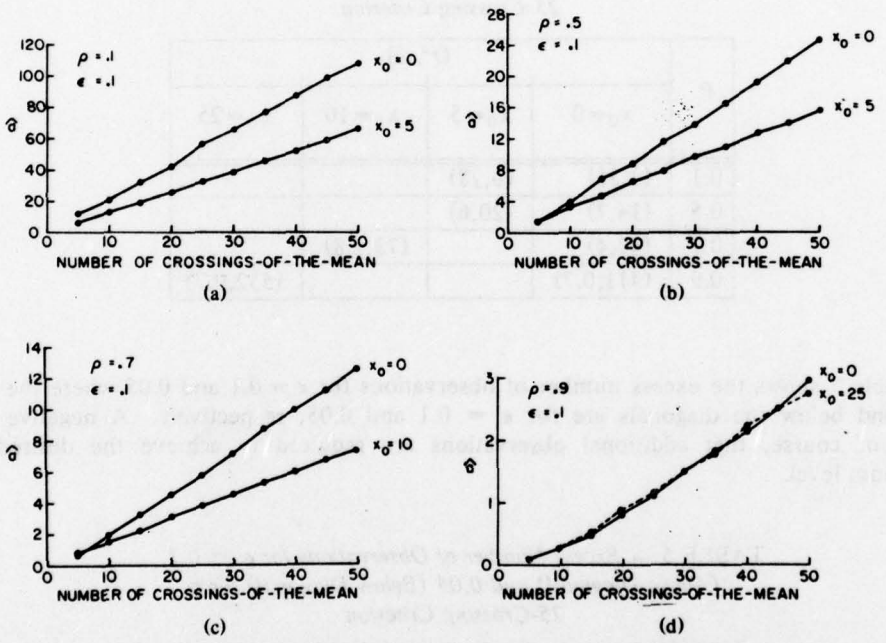


FIGURE 9 — Accuracy characteristics of Crossings-of-the-Mean Rule.

The fact that this rule overestimates t^* is a serious problem in terms of the number of excess observations required beyond the true near-equilibrium point. To illustrate this point, consider Table 3, which shows the pairs $(t^*, \hat{a})$ for the cases tested. The first coordinate of the pair is the true t^* value for $\epsilon = 0.1$ and the second coordinate is the estimated accuracy, $\hat{a}$. Thus, for example, the excess number of observations in the case of $\rho = 0.1$ and $x_0 = 0$ is $(3)(55 - 1) = 162$, and this is also the largest excess. In all situations of excess, the actual attained value of ϵ is nearly 0, i.e., the process is essentially in equilibrium.

Now, if we wanted to be within an $\epsilon = 0.05$ of μ_∞, the $(t^*, \hat{a})$ would be as shown in Table 4. The maximum excess is the same value of 162 but the other excesses have been reduced at the expense of an accuracy of only 0.7 for $\rho = 0.9$.

TABLE 3 — *$(t^*, \hat{a})$ for $\epsilon = 0.1$ and a 25-Crossing Criterion*

ρ	$(t^*, \hat{a})$			
	$x_0=0$	$x_0=5$	$x_0=10$	$x_o=25$
0.1	(3,55)	(5,33)		
0.5	(10,12)	(16,8)		
0.7	(29,6)		(58,3.5)	
0.9	(272,1)			(413,1)

TABLE 4 — *$(t^*, \hat{a})$ for $\epsilon = 0.05$ and a 25-Crossing Criterion*

ρ	$(t^*, \hat{a})$			
	$x_0=0$	$x_0=5$	$x_0=10$	$x_0=25$
0.1	(3,55)	(6,28)		
0.5	(14,7)	(20,6)		
0.7	(42,4)		(73,2.8)	
0.9	(411,0.7)			(552,0.7)

Table 5 shows the excess number of observations for $\epsilon = 0.1$ and 0.05 where the figures above and below the diagonals are for $\epsilon = 0.1$ and 0.05, respectively. A negative excess means, of course, that additional observations are required to achieve the desired near-equilibrium level.

TABLE 5 — *Excess Number of Observations for $\epsilon = 0.1$ (Above Diagonal) and 0.05 (Below Diagonal) for a 25-Crossing Criterion*

ρ	$x_0=0$	$x_0=5$	$x_0=10$	$x_o=25$
0.1	162 / 162	160 / 152		
0.5	110 / 84	112 / 100		
0.7	145 / 126		145 / 131	
0.9	0 / -123			0 / -166

If a 30-crossing criterion is specified, the corresponding results are shown in Tables 6, 7, and 8. In this case, the accuracies are always greater than one and result in an increase in the excess number of observations.

TABLE 6 — $(t^*, \hat{a})$ *for* $\epsilon = 0.1$ *and a 30-Crossing Criterion*

ρ	$(t^*, \hat{a})$			
	$x_0 = 0$	$x_0 = 5$	$x_0 = 10$	$x_0 = 25$
0.1	(3,65)	(5,40)		
0.5	(10,14)	(16,10)		
0.7	(29,7)		(58,4.5)	
0.9	(272,1.5)			(413,1.5)

TABLE 7 — $(t^*, \hat{a})$ *for* $\epsilon = 0.05$ *and a 30-Crossing Criterion*

ρ	$(t^*, \hat{a})$			
	$x_0 = 0$	$x_0 = 5$	$x_0 = 10$	$x_0 = 25$
0.1	(3,65)	(6,33)		
0.5	(14,10)	(20,8)		
0.7	(42,4.8)		(73,3.6)	
0.9	(411,1)			(552,1.1)

TABLE 8 — *Excess Number of observations for* $\epsilon = 0.1$ *(Above Diagonal) and 0.05 (Below Diagonal) for a 30-Crossing Criterion*

ρ	$x_0 = 0$	$x_0 = 5$	$x_0 = 10$	$x_0 = 25$
0.1	192 / 192	195 / 192		
0.5	130 / 126	144 / 140		
0.7	174 / 160		203 / 256	
0.9	136 / 0			207 / 55

Curves similar to the accuracy ones were developed for precision. A sample of one of these is shown in Figure 10. In general, precision improves with decreasing ρ and increasing values of the crossing criterion. A summary of precision estimates for all the cases tested is shown in Table 9 for crossings' criteria of 25 (above the diagonal) and 30 (below the diagonal). These values of precision would appear to be intolerable, in view of the fact that the t^* values, say for $\rho = 0.9$, are between 300 and 500.

4.4 Rule 4 (Cumulative-Mean Rule)

This rule was tested using 10,000 arrivals. Table 10 shows the number of replications (NR) and the utilization ratios, ρ, which were used. The initial conditions, x_0, were 0 and 5 for $\rho = 0.5$ and 0 and 25 for $\rho = 0.9$.

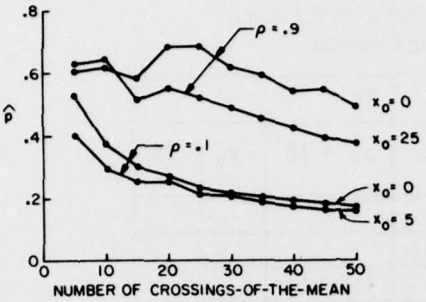


FIGURE 10 — Precision characteristics of the Crossings-of-the-Mean rule.

TABLE 9 — *Precision Values for Crossings' Criteria of 25 (Above Diagonal) and 30 (Below Diagonal)*

ρ	$x_0 = 0$	$x_0 = 5$	$x_0 = 10$	$x_0 = 25$
0.1	0.23 / 0.21	0.22 / 0.21		
0.5	0.28 / 0.24	0.32 / 0.31		
0.7	0.47 / 0.52		0.42 / 0.41	
0.9	0.68 / 0.62			0.52 / 0.48

TABLE 10 — *Utilization Ratios and Number of Replications Used in Testing the Cumulative Mean Rule*

ρ	Number of Replicatons (NR)			
0.5	10	50	100	200
0.9	10	50	100	200

Typical plots of the cumulative mean waiting-time-in-queue versus n (the customer number), for the empty initial condition, are shown in Figures 11a and b. In addition, for $\rho = 0.5$ and NR = 200, a stationary simulation was performed, i.e., one in which the number of customers in the system, when the simulation is begun, is a random variable with the steady-state number in system distribution. The result is shown in Figure 11c. The true steady-state mean delay in queue, in these figures, is one for $\rho = 0.5$ and nine for $\rho = 0.9$.

An inspection of these figures reveals (1) that it is difficult to determine, from a small number of exploratory runs, when stabilization has occurred, (2) that a large number of observations per run are required to make any determination of stability, and (3) that $\hat{t}^*$ is grossly biased positively, even in the stationary case of Figure 11c.

The conclusion seems obvious; the cumulative mean is a very poor method for estimating t^* and we, therefore, do not intend to consider it any further, e.g., no estimate of precision has been obtained since it would require an inordinate amount of computer time. As pointed out by Law [5], it appears to be a technique that would determine the value of n for which

$$\frac{\sum_{i=1}^{n} E[X_i]}{n} \approx \mu_\infty.$$

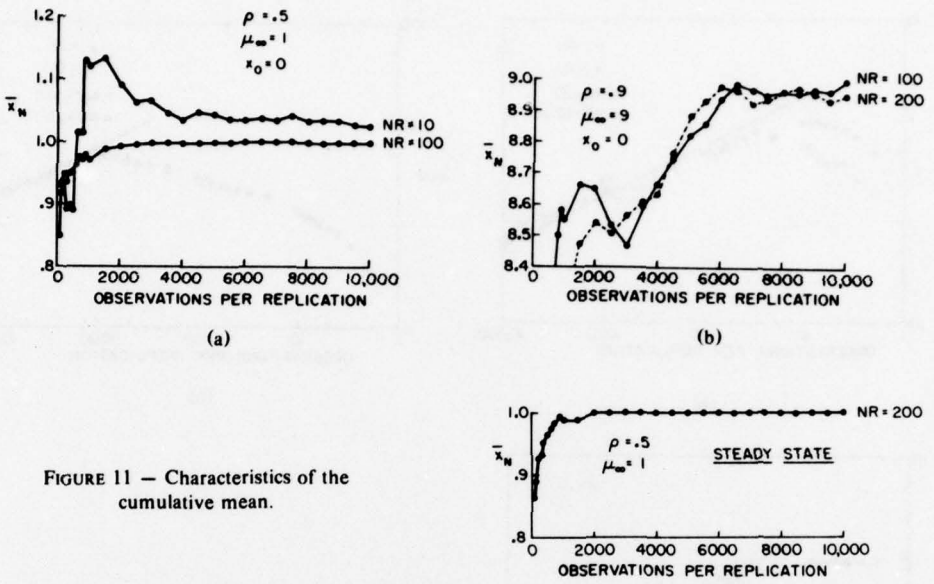


FIGURE 11 — Characteristics of the cumulative mean.

4.5 Rule 5 (Gordon Rule)

This rule was tested using the same conditions as Rule 4 (Cumulative-Mean Rule) and, in fact, the same data. Corresponding to Figures 11a, b, and c are Figures 12a, b, and c. The straight lines on these figures slope downward at the rate of 1 in 2. Again, it is seen that a large number of exploratory runs and observations per run are required in order to carry out the procedure. Also, $\hat{t}^*$ is extremely biased positively, even in the stationary case (Figure 12c). The positive bias is to be expected, since even in the stationary case $\mathrm{Var}[\bar{X}_n]$ is inversely proportional to n only for large values of n. Thus, for $\rho = 0.5$, $\hat{t}^*$ from Figure 12a would be taken as 600; yet the t^* values would be 10 (Table 3) and 14 (Table 4) for $\epsilon = 0.1$ and 0.05, respectively. Similarly, for $\rho = 0.9$, $\hat{t}^*$ would be taken as 1200 (Figure 12b) and the t^* values for $\epsilon = 0.1$ and 0.05 would be 272 (Table 3) and 411 (Table 4), respectively. We conclude, again, that this is a poor method for estimating t^*, and shall not pursue it any further. As in the case of Rule 4, no effort was made to determine its precision characteristics.

5. SUMMARY

In this paper we have developed a comprehensive framework within which may be pursued a study of the problem of the initial transient with respect to mean value. In addition, four rules quoted frequently in the literature, plus a natural variant of one of them, have been evaluated in the $M/M/1$ situation. The principal conclusion is that none of the five rules is satisfactory and that they should not be recommended for use by practitioners.

In summary, the Conway and Modified Conway Rules badly underestimate t^*, they are costly in the sense of wasting data, and there exists no procedure for determining either the number or length of exploratory runs.

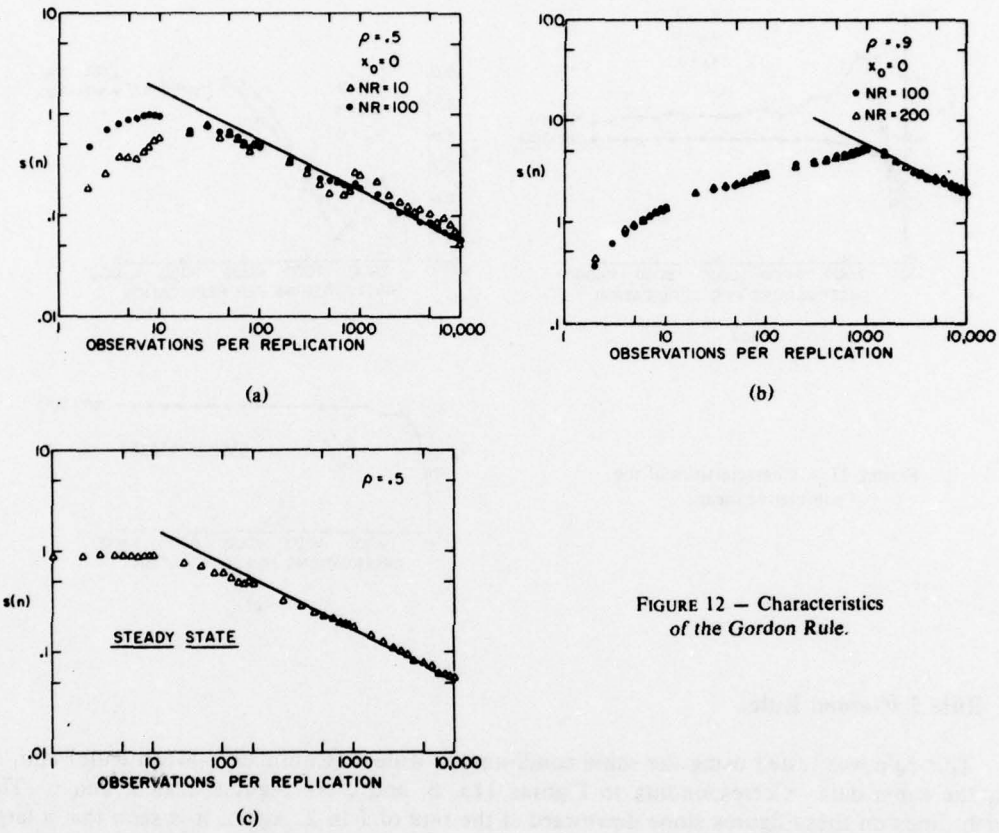


FIGURE 12 — Characteristics of the Gordon Rule.

The Crossings-of-the-Mean Rule when applied to high ρ values, dictates a crossing criterion of about 30. But this causes a large positive bias for the lower ρ values, and hence, is wasteful of data. Also, its precision, especially at the higher ρ values, is unacceptable.

Both the Cumulative-Mean Rule and the Gordon Rule are badly biased positively. They have too many judgmental factors, such as the number and length of replications. If these are not selected to be very large initially, then there is the chance that all the data generated cannot be used, and the whole procedure must be started anew from the beginning; unless, of course, a snapshot of the system at the end of each replication has been preserved, so that additional observations can be added to each run without starting all over. It might be argued that both these techniques, when applied with a sufficient number of runs and observations per run, automatically provide an estimate of the mean. In the case of the Cumulative-Mean Rule, the stabilization value would be that estimate; and in the case of the Gordon Rule, an estimate could be provided by using the cumulative mean of all the observations beyond the truncation point. The point is that both are extremely costly — substantially more data have been generated than are needed to achieve the goal.

REFERENCES

[1] Conway, R.W., "Some Tactical Problems in Digital Simulation," Management Science, *10*, 47-61 (1963).

[2] Fishman, G.S., Concepts and Methoids in Discrete Event Digital Simulation, John Wiley, New York, 1973).

[3] Gordon, G., *System Simulation* (Prentice-Hall, Englewood Cliffs, N.J., 1969).

[4] Heathcote, C.R., and P. Winer, "An Approximation for the Moments of Waiting Times," Operations Research, *17*, 175-186 (1969).

[5] Law, A.M., "A Comparison of Two Techniques for Determining the Accuracy of Simulation Output," Report No. 75-11, Department of Industrial Engineering, University of Wisconsin at Madison, (1975).

[6] Morisaku, T., "Techniques for Data-Truncation in Digital Computer Simulation," Ph.D. Dissertation, Industrial and Systems Engineering Department, University of Southern California, Los Angeles, Calif., (1976).

PROBABILISTIC FORMULATIONS OF THE MULTIFACILITY WEBER PROBLEM

Adel A. Aly

School of Industrial Engineering
University of Oklahoma
Norman, Oklahoma

John A. White

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia

ABSTRACT

The problem considered is to locate one or more new facilities relative to a number of existing facilities when both the locations of the existing facilities, the weights between new facilities, and the weights between new and existing facilities are random variables. The new facilities are to be located such that expected distance traveled is minimized. Euclidean distance measure is considered; both unconstrained and chance-constrained formulations are treated.

1. INTRODUCTION

To date, the study of location problems has been restricted primarily to deterministic formulations. In this paper the effect of random variation on the location decision will be studied. Probabilistic formulations of the multifacility Weber problem are developed and solution procedures are obtained. Applications of the various formulations are cited to motivate an understanding of the contexts in which the formulations apply.

The elements of a facilities-location problem which are treated as random variables are the locations of the existing facilities and the amount of interaction between new and existing facilities. As an illustration of a location problem in which random variation can exist, consider the location of a maintenance department for material-handling equipment in an industrial plant. Maintenance performed is of two types, scheduled and unscheduled. Unscheduled maintenance arises when the material-handling equipment fails and a repairman is dispatched to the job site to perform the necessary repairs. The location of the equipment when it fails is a random variable. Additionally, the number of times a particular piece of equipment fails during a year is a random variable. The determination of the location of the maintenance department based on the random variation involved is typical of the location problems considered.

The deterministic formulation of the multifacility Weber problem is given by

$$\text{P1.} \quad \underset{X_j}{\text{minimize}}\ f(X_1, \ldots, X_n) = \sum_{1 \le j < k \le n} V_{jk}|X_j - X_k| + \sum_{j=1}^{n}\sum_{i=1}^{m} W_{ji}|X_j - P_i|,$$

where

X_j = location of new facility j, $j=1, \ldots, n$,
P_i = location of existing facility i, $i=1, \ldots, m$,
V_{jk} = number of trips per unit time between new facilities j and k, for all $j < k$,
W_{ji} = number of trips per unit time between new facility j and existing facility i, for all j, i,
$|X_k - P_i|$ = Euclidean distance between the points X_j and P_i, and
$f(X_1, \ldots, X_n)$ = total distance traveled per unit time as a function of $X_1, \ldots, X_n$.

In P1, the objective is to determine the locations of the new facilities in order to minimize total distance traveled per unit time. In the subsequent discussion, it is assumed that all new facilities are chained [4].

The treatment of the probabilistic variation of P1 includes the possibility of P_i, V_{jk}, and W_{ji} being random variables. Two types of probabilistic problems are studied. In the first case it is assumed that, for a given realization of P_i, a realization of W_{ji} occurs. Once the location of existing facility i is known, all subsequent trips between new facility j and existing facility i will share the same distance $|X_j - P_i|$, and the weight attached to the trip will be W_{ji}. From a probability point of view, the distance traveled between new facility j and existing facility i is expressed as a product of the random variables W_{ji} and $|X_j - P_i|$. In the second case considered, for each trip included in W_{ji}, the distance between new facility j and existing facility i can be different. There are W_{ji} trips during the planning horizon under investigation, and existing facility i changes its location during this planning horizon independent of the weight W_{ji}. For convenience, let P_{ih} denote the location of existing facility i on trip h. Thus, on trip h, the distance traveled is determined from the value of the random variable P_{ih}. The "weight" or number of trips per unit time is considered to be independent of the location of each existing facility. Thus, the distance traveled can be represented as a random sum of random variables.

To illustrate the first of the two cases considered, let us suppose new warehouses are to be located across the country. The sources of goods shipped to the warehouses and the destinations of goods shipped from the warehouses are not known a priori. However, after the warehouses become operational, the locations of suppliers and customers will become known. The number of shipments per month from the suppliers to the warehouses and from the warehouses to customers is not known exactly, but can be expressed in the form of a probability distribution. Since all shipments from supplier i to warehouse j will be from the point P_i, once the value of P_i becomes known the location problem can be formulated as a function of the sum of the products of the random variables $|X_j - P_i|$ and W_{ji}.

As an illustration of the second case, let us suppose a military hospital is to be located to provide medical treatment for personnel wounded in combat. Patients are brought from the area to the hospital in helicopters. There are m combat areas, and the number of helicoter trips to and from combat area i is a random variable W_i. The location of a wounded soldier in combat area i is a random variable denoted by P_i. Thus, each of the W_i trips can be to a different location in combat area i. In this case, the location problem is formulated as a random sum of random variables.

Research on probabilistic formulations of the Weber problem has included the treatment of the single-facility problem by Cooper [6], who assumed P_i is distributed as a bivariate normal

probability density function, W_i is deterministic, and distance is Euclidean. All random variables he assumed to be independent. Cooper employed a convergent, linear iteration technique to minimize expected cost per unit time. The convergence of the algorithm was established by Katz and Cooper [18]. The same iterative algorithm was employed subsequently by Katz and Cooper [19] in treating both exponential and symmetrical exponential distributions as the probability-density function for P_i.

Hurter and Prawda [16] solved the Euclidean, single-facility location problem when the quantities of service demanded are independent random variables. They formulated the problem as a chance-constrained programming problem, but the constraints were used to bound W_i instead of bounding the cost of transportation, which is a function of the distance. In their analysis, the locations of the existing facilities are assumed to be deterministic when the probabilistic problem is transformed to a deterministic equivalent problem. Hurter and Prawda showed that any existing algorithm for solving the deterministic single-facility problem can be used to solve their chance-constrained problem.

The only previous research on a probabilistic formulation of the multifacility Weber problem appears to be the chance constrained formulation of Seppälä [28]. In his model, V_{jk} and W_{ji} are treated as random variables, but P_i is assumed to be deterministic and Euclidean distances are employed. Seppälä employs a fractile criterion rather than an expected-value criterion. Using the approach developed by Charnes and Cooper [5] to convert the chance constraint to its deterministic equivalent, Seppälä obtains a nonlinear objective function. To solve his model, the CHAPS algorithm developed in [27] is used to convert the nonlinear objective function to a linear objective function augmented by some nonlinear convex constraints. A linear approximation algorithm similar to MAP, introduced by Griffith and Steward [13], is employed to solve the resulting formulation.

2. PRINCIPLES OF CHOICE

In modeling a real-world decision problem, Morris [21] suggests that three alternatives are available. The problem can be modeled as a decision under assumed certainty, a decision under risk, or a decision under uncertainty. The research to date on location problems has concentrated on modeling the problem as a decision under assumed certainty. Thus, deterministic approaches were taken, where either the total travel cost was minimized (minisum criterion), or the maximum travel cost was minimized (minimax criterion). The present research effort concentrates only on modeling location problems as decisions under risk.

In a decision under risk it is assumed that the probability distributions are known for all random variables. Also, a number of alternate principles of choice are possible. Sengupta and Portillo-Cambell [26] suggest four possible optimization criteria under conditions of risk:

(a) expected-value criterion

(b) portfolio criterion

(c) aspiration-level criterion

(d) fractile criterion.

The expected-value criterion involves the determination of the location of the new facilities such that an appropriately defined expected-cost function is minimized. The portfolio criterion seeks the location which minimizes the variance of cost, subject to a constraint on the expected cost produced. An aspiration-level criterion is used when the facilities are to be

located such that the probability of cost being less than some specified value is maximized. The fractile criterion is difficult to express verbally; it involves the location of the new facilities so that one minimizes the value on the cumulative distribution function of total cost which represents the acceptable probability of cost exceeding that value. Mathematically, the fractile criterion can be stated as

$$\text{minimize } z$$

$$\text{subject to: } \Pr[f(\mathbf{x}) \leqslant z] \geqslant \alpha,$$

where $f(\mathbf{x})$ is the total cost resulting from $\mathbf{x}$, the vector of coordinate locations on the new facilities, and z is the cost below which total cost occurs with at least a probability of α. The deterministic equivalent form of the fractile criterion (see Ref. [5]) is written as

$$\text{minimize } E[f(\mathbf{x})] + K\ SD[f(\mathbf{x})],$$

where $E[\ldots]$ and $SD[\ldots]$ are the expected total cost and standard deviation, respectively and k is some number of standard deviations derived from the probability distribution for a given value of α and z.

Depending upon the situation considered and the preferences of the decision maker, additional constraints can be added to the four basic principles of choice. As an illustration, one might wish to minimize the total expected distance traveled between all facilities, with the restriction that the probability of the total distance exceeding 100 miles must be less than 0.10.

The aspiration-level and fractile criteria require that the probability distribution for total cost be known, whereas the expected-value and portfolio criteria require a knowledge of at most the first two moments of the distribution of total cost. Consequently, in order to provide sufficient information to model a specific location problem using the appropriate principle of choice, the probability distribution for total cost will be developed for the situations considered. Then the first two moments will be derived.

Geoffrion [12] discussed the above criteria and the relationships between them. Sengupta and Portillo-Campbell [26] and Hazell [15] supported Geoffrion's observations with computational results. For a critique of the effectiveness and the pros and cons of the described criteria, see Hazell [15].

In the sequel, two criteria are considered. The first is an analogue of the minisum criterion for the deterministic case, where the expected total random cost is minimized. Using the expected-value criterion naturally has a higher risk than using other criteria, but if we assume that the variance of each random location is small and the correlations between the existing locations are relatively small, then the expected total cost over a large time horizon gives the decision maker reasonable low-risk information. If the decision maker is concerned about a realization near the tail of the probability distribution, e.g., in the case of the emergency-facilities location problem, where the maximum random distance is of interest, other criteria may be used to reduce the risk. The expected-value criterion may be considered a special case of the fractile criterion, when $k = 0$.

The second criterion considered in this paper is analogous to the fractile criterion. Basically, it involves minimizing expected total cost subject to probabilistic constraints on acceptable levels of risk, i.e., a chance-constrained programming. The fractile criterion is an analogue of the minimax criterion for the deterministic case. Fried [11] compared the chance-constrained

programming with both the fractile and the portfolio models; he demonstrated the stability of the chance-constrained model relative to the other two models. Also it is a consistent method of treating the utility choices involved in trading-off risk.

In Section 3 expected total cost criteria are used as an unconstrained probabilistic problem, and in Section 4 chance-constrained programming is used in a constrained probabilistic problem. The chance-constrained model will provide a good tool for sensitivity analysis on the acceptable level of risk.

3. UNCONSTRAINED PROBABILISTIC FORMULATION*

In this section, the two problems discussed above are formulated mathematically. The first problem is identified as the case when the expected cost is the product of the random variables P_i and W_{ji}; the second problem is associated with the case when the expected cost is a random sum of random variables.

For the case of the product of the random variables, the expected total cost function is given by

$$\text{P2.}\quad \underset{X_j}{\text{minimize}}\ E[f(X_1, \ldots, X_n)] = E\Big[\sum_{1\leqslant j<k\leqslant n} V_{jk}|X_j - X_k| + \sum_{j=1}^{n}\sum_{i=1}^{m} W_{ji}|X_j - P_i|\Big]$$
$$= \sum_{1\leqslant j<k\leqslant n} E[V_{jk}]|X_j - X_k| + \sum_{j=1}^{n}\sum_{i=1}^{m} E[W_{ji}]E[|X_j - P_i|].$$

For the case of a random sum of random variables the problem of minimizing expected total cost is written as

$$\text{P3.}\quad \underset{X_j}{\text{minimize}}\ E[f(X_1, \ldots, X_n)] = E\Big[\sum_{1\leqslant j<k\leqslant n} V_{jk}|X_j - X_k| + \sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{h=1}^{W_{ji}} |X_j - P_{ih}|\Big]$$
$$= \sum_{1\leqslant j<k\leqslant n} E[V_{jk}]|X_j - X_k| + \sum_{j=1}^{n}\sum_{i=1}^{m} E\left[\sum_{h=1}^{W_{ji}} |X_j - P_{ih}|\right].$$

A comparison of P2 and P3 indicates that the two differ only in the expected value of the second set of summations. However, since the value of the random sum of independent and identically distributed random variables is given by the product of their expected values, then solving P2 is equivalent to solving P3.

*As a notational convenience no distinction is made in the random variable $P_i = (a_i, b_i)$ and the value of the random variable.

Since P2 and P3 are equivalent formulations in expected value, it suffices to treat only one of the cases in detail. In the plane, the Euclidean distance between the points X_j and X_k and the points X_j and P_i can be represented by

$$|X_j - X_k| = [(x_{j1} - x_{k1})^2 + (x_{j2} - x_{k2})^2]^{1/2}$$

and

$$|X_j - P_i| = [(x_{j1} - a_i)^2 + (x_{j2} - b_i)^2]^{1/2}.$$

Hence, P2 can be written as

$$\text{P4. minimize } E[f(X_1, \ldots, X_n)] = \sum_{1 \leqslant j < k \leqslant n} E[V_{jk}][(x_{j1} - x_{k1})^2 + (x_{j2} - x_{k2})^2]^{1/2} + \sum_{j=1}^{n} \sum_{i=1}^{m} E[W_{ji}] E[(x_{j1} - a_i)^2 + (x_{j2} - b)^2]^{1/2},$$

where it is assumed that all random variables P_i are normaliy distributed in $2m$ dimensions. Let $P_{2m} = \{a_1, b_1, a_2, b_2, \ldots, a_m, b_m\}$ be normally distributed in $2m$ dimensions with mean vector A_{2m} and positive definite covariance matrix V_{2m}. Hence the frequency function of P_{2m} is expressed as

$$f_{2m}(P_{2m}) = \frac{1}{(2\pi)^m \sqrt{|V_{2m}|}}\, e^{-1/2(P_{2m} - A_{2m})\, V_{2m}^{-1}(P_{2m} - A_{2m})}.$$

Here it is assumed that the random variables are dependent, i.e., the correlation coefficients are greater than zero for some of the random variables. Since the covariance matrix is positive definite, then there exists a nonsingular $2m \times 2m$ matrix Q_{2m} such that if $P'_{2m} = Q_{2m}^{-1} P_{2m}$, the $P'_{2m} = \{a'_1, b'_1, a'_2, b'_2, \ldots, a'_m, b'_m\}$ are independent. See Ref. [22] for a proof. In this case, the new mean vector is $Q_{2m}^{-1} A_{2m}$ and the variances will equal the corresponding eigenvalues.

In location problems it could be assumed, without loss of generality, that the random variables P_i are independent for all i, i.e., the locations of all existing facilities are considered independent. In this case, for the same i the two random variables a_i, b_i are dependent and their frequency distribution is a bivariate normal distribution. The above transformation Q_2^{-1} will rotate the x and y axes through the acute angle Φ, where

$$\tan 2\Phi = 2\sigma_{a_i b_i} | (\sigma_{a_i}^2 - \sigma_{b_i}^2),$$

and will transform the density function to a bivariate normal distribution with new independent variances equal to the corresponding eigenvalues of the covariance matrix. If $\Phi = (\pi|4)$, then $\sigma_{a_i} = \sigma_{b_i} = \sigma_i$.

Thus in this paper only the independent case is discussed since all other situations can be reduced to it. Let

$a_i \sim N(\mu_{a_i}, \sigma_{a_i}^2)$, for all i, $i = 1, \ldots, m$, and

$b_i \sim N(\mu_{b_i}, \sigma_{b_i}^2)$, for all i, $i = 1, \ldots, m$.

To simplify computations, it is assumed that $\sigma_{a_i} = \sigma_{b_i} = \sigma_i$. No distributional assumptions are required for V_{jk} and W_{ji} since only their expected values are required in P4; we let $E[V_{jk}] = \mu_{jk}$ and $E[W_{ji}] = \bar{\mu}_{ji}$.

In the Appendix, the expected Euclidean distance between the points X_j and P_i is obtained. Substituting the expected Euclidean distance in P4 yields

$$\underset{X_j}{\text{minimize}} \sum_{1 \leqslant j < k \leqslant n} \mu_{jk}[(x_{j1} - x_{k1})^2 + (x_{j2} - x_{k2})^2]^{1/2} + \sqrt{\frac{\pi}{2}} \sum_{j=1}^{n} \sum_{i=1}^{m} \bar{\mu}_{ji} \sigma_i H\left(-\frac{1}{2}, 1, -\frac{\lambda_{ji}^2}{2\sigma_i^2}\right),$$

where H is the confluent hypergeometric function defined by (A.10) (see Appendix) and λ_{ji}^2 is defined as

(1) $$\lambda_{ji}^2 = (x_{j1} - \mu_{a_i})^2 + (x_{j2} - \mu_{b_i})^2.$$

It is easily established that the objective function in P4 is strictly convex [2]. Also, based on the differentiability conditions at the optimum, the necessary conditions can be obtained. Taking the partial derivatives of $E[f(X_1, \ldots, X_n)]$ with respect to all X_j and setting them equal to zero gives

(2) $$\frac{\partial E[f]}{\partial x_{j1}} = 0 = \sum_{\substack{k=1 \\ k \neq j}}^{n} \frac{\hat{\mu}_{jk}(x_{j1} - x_{k1})}{D_{jk}} + \frac{1}{2}\sqrt{\frac{\pi}{2}} \sum_{i=1}^{m} \bar{\mu}_{ji} \left(\frac{x_{j1} - \mu_{a_i}}{\sigma_i}\right) H\left(\frac{1}{2}, 2, -\frac{\lambda_{ji}^2}{2\sigma_i^2}\right),$$

$$j = 1, \ldots, n,$$

and

(3) $$\frac{\partial E[f]}{\partial x_{j2}} = 0 = \sum_{\substack{k=1 \\ k \neq j}}^{n} \frac{\hat{\mu}_{jk}(x_{j2} - x_{k2})}{D_{jk}} + \frac{1}{2}\sqrt{\frac{\pi}{2}} \sum_{i=1}^{m} \bar{\mu}_{ji} \left(\frac{x_{j2} - \mu_{b_i}}{\sigma_i}\right) H\left(\frac{1}{2}, 2, -\frac{\lambda_{ji}^2}{2\sigma_i^2}\right),$$

$$j = 1, \ldots, n,$$

where

$$\hat{\mu}_{jk} = \begin{cases} \mu_{jk}, & k > j \\ \mu_{kj}, & k < j \end{cases}$$

and

(4) $$D_{jk} = [(x_{j1} - x_{k1})^2 + (x_{j2} - x_{k2})^2]^{1/2}, \text{for all } j, k.$$

Unfortunately, if any two new facilities j and k have the same location at any time, then $D_{jk} = 0$ and the partial derivatives in (2) and (3) are undefined. Eyster et al. [8] employed a hyperboloid approximation procedure (HAP) to eliminate this situation. To adopt their approach, a positive constant ϵ is introduced under the square root in D_{jk}; consequently, the partial derivatives always exist. Let $\hat{\lambda}_{jk}$ denote the modified D_{jk}, i.e.,

(5) $$\hat{\lambda}_{jk} = [(x_{j1} - x_{k2})^2 + (x_{j2} - x_{k2})^2 + \epsilon]^{1/2}.$$

When we substitute (5) in (2) and (3) and set the derivatives to zero, the following iterative expressions result:

(6) $$x_{j1}^{(h+1)} = \frac{\displaystyle\sum_{\substack{k=1 \\ k \neq j}}^{n} \frac{\hat{\mu}_{jk} x_{k1}^{(h)}}{\hat{\lambda}_{jk}^{(h)}} + \frac{1}{2}\sqrt{\frac{\pi}{2}} \sum_{i=1}^{m} \frac{\bar{\mu}_{ji}}{\sigma_i} \mu_{a_i} H\left(\frac{1}{2}, 2, -\frac{\lambda_{ji}^{2(h)}}{\sigma_i^2}\right)}{\displaystyle\sum_{\substack{k=1 \\ k \neq j}}^{n} \frac{\hat{\mu}_{jk}}{\hat{\lambda}_{jk}^{(h)}} + \frac{1}{2}\sqrt{\frac{\pi}{2}} \sum_{i=1}^{m} \frac{\bar{\mu}_{ji}}{\sigma_i} H\left(\frac{1}{2}, 2, -\frac{\lambda_{ji}^{2(h)}}{2\sigma_i^2}\right)}$$

and

$$x_{j2}^{(h+1)} = \frac{\displaystyle\sum_{\substack{k=1 \\ k\neq j}}^{n} \frac{\hat{\mu}_{jk} x_{k2}^{(h)}}{\hat{\lambda}_{jk}^{(h)}} + \frac{1}{2}\sqrt{\frac{\pi}{2}} \sum_{i=1}^{m} \frac{\bar{\mu}_{ji}}{\sigma_i} \mu_{b_i} H\left(\frac{1}{2}, 2, -\frac{\lambda_{ji}^{2(h)}}{2\sigma_i^2}\right)}{\displaystyle\sum_{\substack{k=1 \\ k\neq j}}^{n} \frac{\hat{\mu}_{jk}}{\hat{\lambda}_{jk}^{(h)}} + \frac{1}{2}\sqrt{\frac{\pi}{2}} \sum_{i=1}^{m} \frac{\bar{\mu}_{ji}}{\sigma_i} H\left(\frac{1}{2}, 2, -\frac{\lambda_{ji}^{2(h)}}{2\sigma_i^2}\right)} \tag{7}$$

4. CONSTRAINED PROBABILISTIC FORMULATIONS

4.1 Chance Constrained Multifacility Weber Problem: Case I

The first chance-constrained formulation of the multifacility Weber problem considered is that involving products of random variables. The optimization problem is given as

$$\text{P5.} \quad \text{minimize } Z = \sum_{1\leqslant j<k\leqslant n} \mu_{jk} D_{jk} + \sqrt{\frac{\pi}{2}} \sum_{j=1}^{n} \sum_{i=1}^{m} \bar{\mu}_{ji} \sigma_i H\left(-\frac{1}{2}, 1, -\frac{\lambda_{ji}^2}{2\sigma_i^2}\right)$$

$$\text{subject to} \quad \mu_{jk} D_{jk} \leqslant \xi_{jk}, \text{ for all } j, k, \tag{8}$$

$$\text{and} \quad \Pr(W_{ji} R_{ji} \leqslant \xi_{ji}) \geqslant \gamma_{ji}, \quad \begin{matrix} j = 1, \ldots, n, \\ i = 1, \ldots, m, \end{matrix} \tag{9}$$

where

$$D_{jk} = [(x_{j1} - x_{k1})^2 + (x_{j2} - x_{k2})^2]^{1/2}, \quad R_{ji} = [(x_{j1} - a_i)^2 + (x_{j2} - b_i)^2]^{1/2},$$

and ξ_{jk}, ξ_{ji}, and γ_{ji} are known constants. The first set of constraints (8) is deterministic and forms a convex set. Thus, the only probabilistic element in P5 is due to the chance constraints (9). To develop their deterministic equivalent representations, the following approximation is employed: Let the random variable W be normally distributed with mean μ_W and variance σ_W^2. Let the random variable R^2 be defined by $R^2 = (x_1 - a)^2 + (x_2 - b)^2$, where $a \sim N(\mu_a, \sigma^2)$ and $b \sim N(\mu_b, \sigma^2)$. The probability-density function of R^2 can be shown to be [17]

$$g_{R^2}(y) = \frac{1}{2\sigma^2} e^{-\frac{1}{2\sigma^2}(y+\lambda^2)} I_o\left(\frac{\lambda\sqrt{y}}{\sigma^2}\right), \quad 0 < y < \infty, \tag{10}$$

where

$$\lambda^2 = (x_1 - \mu_a)^2 + (x_2 - \mu_b)^2$$

and

I_n = the modified Bessel function of the first kind and of order n.

If we assume that W and R^2 are independent, then the probability density function of the new random variable $Z = WR$ is given by

$$g(z) = \frac{z^{\frac{\nu_1+\nu_2}{2}-1} K_{(\nu_1-\nu_2)/2}(z)}{2^{\frac{\nu_1+\nu_2}{2}-2} \Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}, \quad 0 < z < \infty, \tag{11}$$

where K_ν (ax) denotes the modified Bessel function of the second kind and order ν, and ν_1 and ν_2 denote the degrees of freedom for noncentral chi-square distributed random variables.

To motivate the approximate deterministic equivalent, let $Y = W^2R^2$, where R^2 is distributed as a noncentral $\chi_2^2(\lambda^2)$ with two degrees of freedom and noncentrality parameter λ^2, and W^2 is a noncentral $\chi_1^2(\mu^2)$ with one degree of freedom and noncentrality parameter μ^2 [17]. If we use Patnaik's noncentral chi-square approximation [23], two different χ_ν^2 distributions with degrees of freedom ν_1 and ν_2, respectively, are obtained.

The Mellin transform of χ_ν^2 is given by Webb [30] as*

$$M(f(\chi_{\nu_1}^2 \mid s) = 2^{(s-1)} \left[\frac{\Gamma\left(\frac{\nu_1}{2} + s - 1\right)}{\Gamma\left(\frac{\nu_1}{2}\right)} \right]. \tag{12}$$

Since the Mellin transform of the product of random variables is given by the product of their Mellin transforms, then

$$M(g(y) \mid s) = M(f(\chi_{\nu_1}^2) \mid s) \cdot M(f(\chi_{\nu_2}^2) \mid s) \tag{13}$$

$$= 2^{2(s-1)} \left[\frac{\Gamma\left(\frac{\nu_1}{2} + s - 1\right) \Gamma\left(\frac{\nu_2}{2} + s - 1\right)}{\Gamma\left(\frac{\nu_1}{2}\right) \cdot \Gamma\left(\frac{\nu_2}{2}\right)} \right].$$

In order to find $g(y)$, the density function of y, the inverse Mellin transform of (13) must be obtained:

$$M^{-1} = g(y) = \frac{1}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^{-s} 2^{2(s-1)} \Gamma\left(\frac{\nu_1}{2} + s - 1\right) \Gamma\left(\frac{\nu_2}{2} + x - 1\right) ds. \tag{14}$$

Equation 14 can be expressed alternatively, in order that the inverse will be recognized easily, from the tables of inverse Mellin transform. In (14), let $s' = 2s + \frac{\nu_1}{2} + \frac{\nu_2}{2} - 2$; then $ds = \frac{1}{2} ds'$ and (14) can be written as

$$g(y) = \frac{2^{1-\left(\frac{\nu_1}{2} + \frac{\nu_2}{2}\right)} y^{\frac{\nu_1}{4} + \frac{\nu_2}{4} - 1}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} (\sqrt{y})^{s'} s^{s'-2} \tag{15}$$

$$\Gamma\left(\frac{s'}{2} + \frac{\nu_2}{4} - \frac{\nu_2}{4}\right) \Gamma\left(\frac{s'}{2} + \frac{\nu_2}{4} - \frac{\nu_1}{4}\right) ds'.$$

*The Mellin transform of the random variable χ is defined as $E(\chi^{s-1})$.

In Ref. [7], on p. 331, the following Mellin transform is given:

$$K_\nu(ax) = M^{-1}\left[a^{-s}\,2^{s-2}\,\Gamma\left(\frac{s}{2}-\frac{\nu}{2}\right)\Gamma\left(\frac{s}{2}+\frac{\nu}{2}\right)\right], \tag{16}$$

where $k_\nu(ax)$ is the modified Bessel function of the second kind and order ν. Observing the similarity between (15) and (16), we conclude that the probability density function for $Y = W^2R^2$ is

$$g(y) = \frac{y^{\frac{\nu_1}{4}+\frac{\nu_2}{4}-1}}{2^{\frac{\nu_1}{2}+\frac{\nu_2}{2}-1}\,\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\,K_{\left(\frac{\nu_1-\nu_2}{2}\right)}\left(y^{1/2}\right),\quad 0 < y < \infty. \tag{17}$$

To simplify (17), let $y^{1/2} = z$, then $2zdz = dy$. Thus, the probability density function for $Z = WR$ can be written

$$g(z) = \frac{z^{\frac{\nu_1+\nu_2}{2}-1}\,K_{(\nu_1-\nu_2)/2}(z)}{2^{\frac{\nu_1+\nu_2}{2}-2}\,\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)},\quad 0 < z < \infty,$$

which has the desired form.

If the cumulative distribution of $g(z)$ is required, some assumptions are made first to obtain a closed form for the integral of (11). It is first assumed that all weights (W_i) are such that $0 \leqslant W_{ji} < 1$; this is achieved by dividing each weight by $\sum_{j=1}^{n}\sum_{i=1}^{m} W_i + \sum_{1\leqslant j<k\leqslant n} V_{jk}$. Thus, the weights can then be defined as the fraction of the total weight. Patnaik [23] provided the following approximation to a noncentral χ^2 with degrees of freedom ν and noncentrality parameter λ:

$$\chi'^2_\nu(\lambda^2) = c\chi^2_f \tag{18}$$

where

$$c = \frac{\nu+2\lambda}{\nu+\lambda} \quad \text{and} \quad f = \nu + \frac{\lambda^2}{\nu+2\lambda}.$$

Therefore, when the distribution of W_i^2, which is $\chi'^2_1(\mu_i^2)$ distributed, is approximated by a central χ^2_f, the degrees of freedom are

$$f = 1 + \frac{\mu^4}{1+2\mu^2}. \tag{19}$$

From the assumption that the random variable W_i takes values below one, the second term in (19) is always a fraction.

Given two chi-square distributions, each with degrees of freedoms n_1, n_2, respectively, if $n_1 > n_2$, then

$$\Pr\left(\chi^2_{n_1} \leqslant \xi^2\right) \leqslant \Pr\left(\chi^2_{n_2} \leqslant \xi^2\right). \tag{20}$$

It may be seen that using n_1 instead of n_2 will underestimate the probability. Hence, there is no overestimation in the probability in (9) if we assume that f defined in (19) equals two, since the difference in this range is small for the χ^2 distribution. Since

$$\Pr\,(W_{ji}R_{ji} \leqslant \xi_{ji}) = \Pr\,(W_{ji}^2\,R_{ji}^2 \leqslant \xi_{ji}^2), \tag{21}$$

applying Patnaik's approximation [$\chi'^2_\nu(\lambda^2) = c\chi^2_f$] yields

$$\Pr\,(W_{ji}^2\,R_{ji}^2 \leqslant \xi_{ji}^2) \simeq \Pr(W_{ji}^2\,R_{ji}^2 \leqslant \xi_{ji}^2/c_1c_2\bar{\sigma}_{ji}^2),$$

where c_1 and c_2 are defined as in (18).

Under the assumption that $\nu_2 = 2$, (11) reduces to

$$g(z) = \frac{z^{\frac{\nu_1}{2}}}{2^{\left[\frac{\nu_1}{2}-1\right]}\Gamma\left(\frac{\nu_1}{2}\right)}\,K_{(\nu_1-2)/2}(z), \qquad 0 < z < \infty. \tag{22}$$

The cumulative distribution is obtained by integrating (22) over z. Letting

$$F(\alpha_{ji}) = \Pr\,(W_{ji}\,R_{ji} \leqslant \alpha_{ji}), \tag{23}$$

where $\alpha_{ji}^2 = \xi_{ji}^2/c_1c_2\bar{\sigma}_{ji}^2$, and noting that

$$\int_0^w z^{n+1}\,K_n(z)\,dz = -w^{(n+1)}\,K_{n+1}(w) + 2^n\,\Gamma(n+1), \tag{24}$$

substitution of (24) in (25) yields

$$F(\alpha_{ji}) = 1 - \frac{(\alpha_{ji})^{\frac{\nu_1}{2}}\,K_{\frac{\nu_1}{2}}(\alpha_{ji})}{2^{\left[\frac{\nu_1}{2}-1\right]}\Gamma\left(\frac{\nu_1}{2}\right)}. \tag{25}$$

Therefore, the chance constraints in P5 can be written as

$$1 - \gamma_{ji} \geqslant \frac{\alpha_{ji}^{\frac{\nu_1}{2}}\,K_{\nu_1/2}(\alpha_{ji})}{2^{\left[\frac{\nu_1}{2}-1\right]}\Gamma\left(\frac{\nu_1}{2}\right)}, \quad \text{for all } j = 1, \ldots, n \text{ and } i = 1, \ldots, m \tag{26}$$

Note that α_{ji} is a function of c_1, and from (18) it is clear that it is a function of X_j; in the same manner ν_1 is a function of X_j. Since $K_\nu(\cdot)$ is well tabulated and available for computer calculations, an iterative method to solve the nonlinear programming problem is recommended. Problem P5 may be stated in a deterministic equivalent form as

$$\text{P6.}\quad \underset{X_j}{\text{minimize}}\ Z = \sum_{1 \leqslant j < k \leqslant n} \mu_{jk}D_{jk} + \sqrt{\frac{\pi}{2}}\sum_{j=1}^{n}\sum_{i=1}^{m}\bar{\mu}_{ji}\sigma_i\,H\left(-\frac{1}{2},\,1,\,-\frac{\lambda_{ji}^2}{2\sigma_i^2}\right)$$

$$\text{subject to}\ \ \mu_{jk}D_{jk} \leqslant \xi_{jk}, \quad \text{for all } j,\ k,$$

$$\text{and} \quad \frac{\alpha_{ji}^{\frac{\nu_1}{2}} K_{\nu_{1/2}}(\alpha_{ji})}{2^{\left(\frac{\nu_1}{2}-1\right)} \Gamma\left(\frac{\nu_1}{2}\right)} \leqslant 1 - \gamma_{ji}, \quad \text{for all } j,i.$$

For a constant ν_1, it is easily shown that the last set of constraints forms a convex set (from the definition of $K(\cdot)$); the first set of constraints also forms a convex set. Thus, the joint constraint set is convex. It can be shown that the objective function is strictly convex; thus, if a local optimum is achieved, it is also a global optimum. Hence, a number of nonlinear programming algorithms may be employed to obtain the solution to P6. In the case that ν_1 is not considered as a parameter, i.e., ν_1 is a function of X_j, the convexity condition of the constraints may not hold. However, a local optimum solution is still available and it may turn out to be a global optimum solution.

It is easier to work with the constraints when ν_1 is treated as a parameter since $K_{\nu_{1/2}}(\cdot)$ will have the same order during all iterations. However, this may be accomplished, without loss of generality, if the known (μ_{a_i}, μ_{b_i}) are rescaled so that any coordinate takes a value between zero and one. This will imply that λ_{ji} is bounded as $0 \leqslant \lambda_{ji} \leqslant 1$, from (8), $\nu_1 \approx 2$ and the constraints given by (26) are written as

$$\alpha_{ji} K_1(\alpha_{ji}) \leqslant 1 - \gamma_{ji}, \quad \text{for all } i,j,$$

which will simplify the computations dramatically. Note that the optimum solution has to be adjusted to its former scale so that the total cost obtained is in the correct units.

4.2 Chance-Constrained Multifacility Weber Problem: Case II

The second type of problem to be considered involves a random sum of random variables. The optimization problem is formulated as

$$\text{P7.} \quad \underset{X_j}{\text{minimize}}\ Z = \sum_{1 \leqslant j < k \leqslant n} \mu_{jk} D_{jk} + \sqrt{\frac{\pi}{2}} \sum_{j=1}^{n} \sum_{i=1}^{m} \bar{\mu}_{ji} \sigma_i H\left(-\frac{1}{2}, 1, -\frac{\lambda_{ji}^2}{2\sigma_i^2}\right)$$

$$\text{subject to} \quad \mu_{jk} D_{jk} \leqslant \xi_{jk}, \quad \text{for all } j,k$$

$$\text{and} \quad \Pr\left[\sum_{h=1}^{W_{ji}} R_{ji}^{(h)} \leqslant \xi_{ji}\right] \geqslant \gamma_{ji}, \quad \begin{matrix} j = 1, \ldots, n, \\ i = 1, \ldots, m, \end{matrix}$$

where D_{jk}, R_{ji}, ξ_{ji}, γ_{ji}, H, and λ_{ji} are as defined previously. In order to solve P7 the chance constraints are converted to equivalent deterministic constraints. If we define the random variable Y_{ji} as

$$Y_{ji} = \sum_{h=1}^{W_{ji}} R_{ji}^{(h)},$$

we may conclude, under very general conditions, that Y_{ji} is approximately normally distributed with mean

$$E[Y_{ji}] = E[W_{ji}]E[R_{ji}] \tag{27}$$

and variance

$$V[Y_{ji}] = E[W_{ji}]V[R_{ji}] + V[W_{ji}]E^2[R_{ji}]. \tag{28}$$

From (A.12),

$$E[R_{ji}] = \sqrt{\frac{\pi}{2}}\,\sigma_i\, H\left(-\frac{1}{2},\ 1,\ -\frac{\lambda_{ji}^2}{2\sigma_i^2}\right) \tag{29}$$

and

$$E[R_{ji}^2] = 2\sigma_i^2 + \lambda_{ji}^2. \tag{30}$$

For a detailed discussion of the conditions underlying the Central Limit Theorem for the sum of a random number of independent random variables, see Blum, *et al* [3] and Rényi [24], [25].

The chance constraints can be written in normalized form as

$$\Pr\left(\frac{Y_{ji} - E[Y_{ji}]}{SD[Y_{ji}]} \leqslant \frac{\xi_{ji} - E[Y_{ji}]}{SD[Y_{ji}]}\right) \geqslant \gamma_{ji} \quad \text{for all } j,i, \tag{31}$$

where $SD[Y_{ji}]$ denotes the standard deviation of the random variable Y_{ji}. Equivalently, (31) can be expressed as

$$\xi_{ji} \geqslant E[Y_{ji}] + SD[Y_{ji}]\phi^{-1}(\gamma_{ji}), \quad \begin{matrix} j = 1, \ldots, n, \\ i = , \ldots, m. \end{matrix} \tag{32}$$

The deterministic equivalent of P7 is given by

$$\text{P8.}\quad \underset{X_j}{\text{minimize}}\ z = \sum_{1 \leqslant j < k \leqslant n} \mu_{jk} D_{jk} + \sqrt{\frac{\pi}{2}} \sum_{j=1}^{n} \sum_{i=1}^{m} \bar{\mu}_{ji}\sigma_i\, H\left(-\frac{1}{2},\ 1,\ -\frac{\lambda_{ji}^2}{2\sigma_i^2}\right)$$

$$\text{subject to}\ \ \mu_{jk} D_{jk} \leqslant \xi_{jk},\ 1 \leqslant j < k \leqslant n,$$

$$\text{and}\ \ E[Y_{ji}] + SD[Y_{ji}]\phi^{-1}(\gamma_{ji}) \leqslant \xi_{ji}, \quad \begin{matrix} j = 1, \ldots, n, \\ i = 1, \ldots, m. \end{matrix}$$

It can be shown that $E[Y_{ji}]$ is a convex function and the objective function is strictly convex [2]. However, the convexity of $SD[Y_{ij}]$ may not hold. Hence, a local optimum solution is guaranteed using any convergent convex programming algorithm.

APPENDIX

1. Derivation of $E[(x_1-a)^2 + (x_2-b)^2]$:

Let $R^2 = (x_1-a)^2 + (x_2-b)^2$. From (10)

$$g_{R^2}(y) = \frac{1}{2\sigma^2}\, e^{-\frac{1}{2\sigma^2}(y+\lambda^2)}\, I_0\left(\frac{\lambda\sqrt{y}}{\sigma^2}\right),\ 0 < y < \infty. \tag{A.1}$$

The expected value of R^2 is derived as follows:

$$E[R^2] = \int_0^\infty y\, g_{R^2}(y)\, dy$$

$$= \int_o^\infty \frac{y}{2\sigma^2}\, e^{-\frac{1}{2\sigma^2}(y+\lambda^2)}\, I_o\left[\frac{\lambda\sqrt{y}}{\sigma^2}\right] dy$$

Let $\bar{\lambda}^2 = \dfrac{\lambda^2}{2\sigma^2}$, $\dfrac{y}{2\sigma^2} = w$, then $dw = \dfrac{1}{2\sigma^2}\, dy$

$$E[R^2] = 2\sigma^2\, e^{-\bar{\lambda}^2} \int_0^\infty we^{-w}\, I_o(2\bar{\lambda}\sqrt{w})\, dw \tag{A.2}$$

But from Ref. [1], p. 375,

$$I_o(z) = \sum_{k=0}^{\infty} \frac{\left[\frac{z^2}{4}\right]^k}{(k!)^2} = \sum_{k=0}^{\infty} \frac{\left[\frac{z}{2}\right]^{2k}}{(k!)^2} \tag{A.3}$$

Substituting (A.3) in (A.2) yields

$$E[R^2] = 2\sigma^2\, e^{-\bar{\lambda}^2} \sum_{k=0}^{\infty} \frac{(\bar{\lambda})^{2k}}{(k!)^2} \int_0^\infty we^{-w}\, (\sqrt{w})^{2k}\, dw.$$

It can be seen that the integral is a gamma function where

$$\Gamma(z) = \int_0^\infty t^{z-1}\, e^{-t}\, dt$$

from Ref. [1], p. 255. Therefore, the integral equals

$$\int_0^\infty w^{k+1}\, e^{-w}\, dw = \Gamma(k+2) = (k+1)(k!) \tag{A.4}$$

Substituting (A.4) for the integral, we obtain

$$E[R^2] = 2\sigma^2\, e^{-\bar{\lambda}^2} \sum_{k=0}^{\infty} \frac{(\bar{\lambda}^2)^k}{k!} \cdot (k+1)$$

$$= 2\sigma^2\, e^{-\bar{\lambda}^2} \left[\sum_{k=0}^{\infty} \frac{(\bar{\lambda}^2)^k}{k!} + \sum_{k=1}^{\infty} \frac{k(\bar{\lambda}^2)^k}{k!}\right].$$

Since $e^z = \sum_{k=0}^{\infty} \dfrac{(z)^k}{k!}$,

$$E[R^2] = 2\sigma^2\, e^{-\bar{\lambda}^2} \left[e^{\bar{\lambda}^2} + (\bar{\lambda}^2)\, e^{\bar{\lambda}^2}\right]$$

$$= 2\sigma^2 + 2\sigma^2\, \bar{\lambda}^2.$$

Since $\bar{\lambda}^2 = \dfrac{\lambda^2}{2\sigma^2}$,

$$E[R^2] = 2\sigma^2 + \lambda^2. \tag{A.5}$$

2. Derivation of $E\left[[(x-a_i)^2 + (y-b_i)^2]^{1/2}\right]$:

Given that the random variable P behaves according to the normal distribution, i.e., $a \sim N(\mu_a, \sigma^2)$, $b \sim N(\mu_b, \sigma^2)$, we let R denote the statistic given by $R = [(x_1-a)^2 + (x_2-b)^2]^{1/2}$. Then the probability density function of R is

$$\bar{g}_R(r) = \frac{r}{\sigma^2}\, e^{-\frac{1}{2\sigma^2}(r^2+\lambda^2)}\, I_o\left(\frac{\lambda r}{\sigma^2}\right),\ 0 < y < \infty, \tag{A.6}$$

where

$\lambda^2 = (x_1-\mu_a)^2 + (x_2-\mu_b)^2$, and

I_n = the modified Bessel function of the first kind and order n.

The probability density function of R^2 is given in (15). Since the statistic R equals the square root of the statistic R^2, then

$$\bar{g}_R(r) = g_R(\sqrt{y}) \cdot \left|\frac{dy}{dr}\right|,$$

where $r = \sqrt{y}$; thus $2rdr = dy$ and the Jacobian $\left|\frac{dy}{dr}\right| = 2r$. Therefore,

$$\bar{g}_R(r) = 2r\, \frac{1}{2\sigma^2}\, e^{-\frac{1}{2\sigma^2}(r^2+\lambda^2)}\, I_o\left(\frac{\lambda r}{\sigma^2}\right),\ 0 < y < \infty,$$

which gives the probability density function shown in (A.6) Given the probability density function of R, we derive the expected value as

$$E[R] = \int_0^\infty r\, \bar{g}_R(r)\, dr$$

$$= \int_0^\infty \frac{r^2}{\sigma^2}\, e^{-\frac{1}{2\sigma^2}(r^2+\lambda^2)}\, I_o\left(\frac{\lambda r}{\sigma^2}\right) dr.$$

Using the expansion of $I_o(z)$ given in (A.3),

$$E[R] = \frac{e^{-\frac{\lambda^2}{2\sigma^2}}}{\sigma^2} \int_0^\infty r^2 \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda r}{2\sigma^2}\right)^{2k}}{(k!)^2}\, e^{-\frac{r^2}{2\sigma^2}}\, dr.$$

Let $\frac{r^2}{2\sigma^2} = w$, then $r = \sqrt{2}\sigma\sqrt{w}$, $dr = \frac{\sigma}{\sqrt{2}}\, w^{-1/2} dw$, giving

$$E[R] = \frac{e^{-\frac{\lambda^2}{2\sigma^2}}}{\sigma^2} \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda}{2\sigma^2}\right)^{2k}}{(k!)^2} \int_0^\infty \left(\sqrt{2}\,\sigma\right)^{2k+2} w^{k+1}\, e^{-w}\, \frac{\sigma}{\sqrt{2}}\, w^{-\frac{1}{2}}\, dw$$

$$= \sqrt{2}\sigma\, e^{-\frac{\lambda^2}{2\sigma^2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda}{\sqrt{2}\sigma}\right)^{2k}}{(k!)^2} \int_0^\infty w^{k+\frac{1}{2}}\, e^{-w}\, dw. \tag{A.7}$$

But the integral is a gamma function [1], p. 255, thus

$$\int_0^\infty w^{k+\frac{1}{2}} e^{-w}\, dw = \Gamma\left(k+\frac{3}{2}\right). \tag{A.8}$$

Substituting the value of the integral in (A.7), we obtain

$$E[R] = \sqrt{2}\,\sigma\, e^{-\frac{\lambda^2}{2\sigma^2}} \sum_{k=0}^{\infty} \frac{\left(\frac{\lambda^2}{2\sigma^2}\right)^k}{k!} \frac{\Gamma\left(k+\frac{3}{2}\right)}{\Gamma(k+1)} \cdot \frac{\Gamma(1)}{\Gamma\left(\frac{3}{2}\right)} \cdot \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma(1)} \tag{A.9}$$

From Ref. [1], p. 504,

$$H(a,b,z) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)}{\Gamma(a)} \frac{\Gamma(b)}{\Gamma(b+k)} \frac{z^k}{k!}. \tag{A.10}$$

Substituting (A.10) in (A.9) we obtain

$$E[R] = \sqrt{2}\,\sigma\,\Gamma\left(\frac{3}{2}\right) e^{-\frac{\lambda^2}{2\sigma^2}} H\left(\frac{3}{2},\, 1,\, \frac{\lambda^2}{2\sigma^2}\right).$$

Since $\Gamma\left(\frac{3}{2}\right) = \sqrt{\frac{\pi}{2}}$,

$$E[R] = \sqrt{\frac{\pi}{2}}\,\sigma\, e^{-\frac{\lambda^2}{2\sigma^2}} H\left(\frac{3}{2},\, 1,\, \frac{\lambda^2}{2\sigma^2}\right). \tag{A.11}$$

Equation (A.11) may be further simplified by using the Kummer Transformation (Ref. [1], p. 505) which is stated as

$$H(a,b,z) = e^z\, H(b-a,b,-z)$$

Application of the Kummer Transformation to (A.11) yields the following expected value of R,

$$E[R] = \sqrt{\frac{\pi}{2}}\,\sigma\, H\left(-\frac{1}{2},\, 1,\, -\frac{\lambda^2}{2\sigma^2}\right) \tag{A.12}$$

BIBLIOGRAPHY

[1] Abramowitz, M., and I.A. Stegun, *Handbook of Mathematical Functions,* (Dover, New York, 1968).

[2] Aly, A.A., "Probabilistic Formulations of Some Facility Location Problems," Unpublished Ph.D. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg (1975).

[3] Blum, J.R., D.L. Hanson, and J.I. Rosenblatt, "On the Central Limit Theorem for the Sum of a Random Number of Independent Random Variables," Zeitschift fur Wahrscheinlichkeitstheorie und verwandt Gebiete, *1*, (1963) 389-393.

[4] Cabot, A.V., R.L. Francis, and M.A. Stary, "A Network Flow Solution to the Rectilinear Distance Facility Location Problem," AIIE Transactions, *2*, 132-141 (1970).

[5] Charnes, A., and W.W. Cooper, "Deterministic Equivalents for Optimizing and Satisfying Under Chance Constrain," Operations Research, *11*, 18-39 (1963).

[6] Cooper, Leon, "A Random Locational Equilibrium Problem," Journal of Regional Science, *14*, 47-54 (1974).

[7] Erdelyi, A., W. Mangus, F. Oberhettinger, and F.G. Tricomi, *Tables of Integral Transforms*, Vol. 1, Bateman Manuscript Project (McGraw-Hill, New York, 1954).

[8] Eyster, J.W., J. A. White, and W.W. Wierwille, "On Solving Multifacility Location Problems Using a Hyberboloid Approximation Procedure," AIIE Transactions, *5*, 1-6 (1973).

[9] Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. 1 (John Wiley, New York, 1968).

[10] Francis, R.L., and J.A. White, *Facilities Layout and Location: An Analytical Approach*, (Prentice-Hall, Englewood Cliffs, N.J., 1974).

[11] Fried, Joel, "Bank Portfolio Selection," Journal of Finance and Quantitative Analysis, *5*, 203-227 (1970).

[12] Geoffrion, A.M., "Stochastic Programming with Aspiration on Fractile Criteria, Management Science, *13*, 672-679 (1967).

[13] Griffith, R.E., and R.A. Stewart, "A Nonlinear Programming Technique for Optimization of Continuous Processing Systems," Management Science, *7*, 379 (1961).

[14] Hadley, G., *Nonlinear and Dynamic Programming*, (Addison-Wesley, Reading, Mass. 1964).

[15] Hazell, P.B.R., "Comment on the Fractile Approach to Linear Programming Under Risk," Management Science, *17*, 236-237 (1970).

[16] Hurter, A.P., Jr., and J. Prawda, "A Warehouse Location Problem with Probabilistic Demand," Working Paper Series Number 42, Graduate School of Business Administration, Tulane University, New Orleans, Louisiana.

[17] Johnson, N.L. and S. Kotz, *Continuous Univariate Distributions*, Vol. 2 (Houghton-Mifflin Boston, 1970).

[18] Katz, Norman, and Leon Cooper, "An Always-Convergent Numerical Scheme for a Random Locational Equilibrium Problem," SIAM Journal of Numerical Analysis, *11*, 683-691 (1974).

[19] Katz, J.N., and L. Cooper, "Normally and Exponentially Distributed Locational Equilibrium Problems," Journal of Research of the National Bureau of Standards, *80B*, 53-73 (1976).

[20] Kuhn, H.W., "One Pair of Dual Nonlinear Problems," in *Nonlinear Programming*, Chapter 3, J. Abadie, ed., (John Wiley, New York, 1967).

[21] Morris, W.T., "On the Art of Modeling," Management Science, *13*, 707-717 (1967).

[22] Morrison, D.F., *Multivariate Statistical Methods*, (McGraw-Hill, New York, 1976).

[23] Patnaik, P.B., "The Non-Central χ^2 and F Distributions and Their Applications," Biometrika, *36*, 202-232 (1949).

[24] Renyi, A., "On the Asymptotic Distribution of the Sum of a Random Number of Independent Random Variables," Acta Mathematica, *8*, 193-199 (1957).

[25] Renyi, A., "On the Central Limit Theorem for the Sum of a Random Number of Independent Random Variables," Acta Mathematica, *11*, 97-102 (1960).

[26] Sengupta, J.K., and J.H. Portillo-Campbell, "A Fractile Approach to Linear Programming Under Risk," Management Science, *16*, 298-308 (1970).

[27] Seppala, Y., "Constructing Sets of Uniformly Tighter Linear Approximation for a Chance Constraint," Management Science, *17*, 736-749 (1971).

[28] Seppala, Y., "On a Stochastic Multi-Facility Location Problem," AIIE Transactions, *7*, 56-62 (1975).

[29] Shanno, D.F. and R.L. Weil, "Linear Programming with Absolute-Value Functionals," Operations Research, *19*, 120-124 (1971).

[30] Webb, L.R., "On the Distribution of the Product of Diode Detector Waveforms," Canadian Journal of Physics, *34*, 679-691 (1956).

THE CONSTRAINED SHORTEST PATH PROBLEM

Y.P. Aneja and K.P.K. Nair

University of New Brunswick
Fredericton, N.B., Canada

ABSTRACT

The shortest path problem between two specified nodes in a general network possesses the unimodularity property and, therefore, can be solved by efficient labelling algorithms. However, the introduction of an additional linear constraint would, in general, destroy this property and the existing algorithms are not applicable in this case. This paper presents a parametric approach for solving this problem. The algorithm presented would require, on the average, a number of iterations which is polynomially bounded. The similarity of this approach to that of the generalized Lagrange multiplier technique is demonstrated and a numerical example is presented.

1. INTRODUCTION

The problem of determining the shortest chain between a pair of specified nodes in a general network is of interest in many ways [5]. The problem is formulated as a special case of the minimal-cost-flow problem [4], and efficient algorithms are available for computing the shortest path. The shortest path so obtained will minimize a particular linear attribute (function) of the path such as cost, time, or distance. In the constrained shortest path problem, the optimal path must adhere to an additional linear constraint and minimize the chosen attribute. A typical formulation of this problem will be one of finding a path with minimum distance subject to a budgetary constraint. Alternatively, it may be one of minimizing the cost subject to a constraint on time. The introduction of such an additional constraint destroys the unimodularity of the constraint matrix, and any simplex-based algorithm would not, in general, guarantee an integral solution.

In this paper, we treat the additional constraint also as an objective and formulate a bicriteria linear program. The optimal solution to the original problem is shown to be a special kind of extreme point of the bicriteria problem. An algorithm is presented for obtaining such an extreme point. At each iteration of the algorithm, a shortest path that minimizes a positively weighted average of the two objectives is determined. It is shown that one needs to solve, on the average, at most n such problems, where n is the number of variables, to obtain the desired nondominated extreme point. It is of interest to see that the process has strong similarity to the generalized Lagrange multiplier technique [3], and this aspect is demonstrated. A numerical example illustrates the algorithm.

2. FORMATION OF THE PROBLEM

Let $[N; A]$ be a network having a single source s and a single sink t. Let $a(x, y)$ be the cost of transporting a unit flow along the arc (x, y), and let $b(x, y)$ denote the traverse time along this arc. The constrained shortest path problem is formulated as:

$$\text{Minimize } z_1(f) = \sum_A a(x, y) \cdot f(x, y)$$

$$\text{subject to } \sum_{y \in N} f(x, y) - \sum_{y \in N} f(y, x) = \begin{cases} 1 \text{ if } x = s, \\ -1 \text{ if } x = t, \\ 0 \text{ otherwise,} \end{cases}$$

(P1) $$\text{and } \sum_A b(x,y) \cdot f(x, y) \leqslant B,$$

$$f(x, y) = 0 \text{ or } 1.$$

It is assumed that $a(x, y)$ and $b(x, y)$ are nonnegative for all $(x, y) \in A$, and the vector b (with $b(x, y)$ as its components) is not a multiple of a.

Consider, now, the following bicriteria linear program.

$$\text{Minimize } z_1(f) = \Sigma\, a(x, y) \cdot f(x, y)$$

$$\text{and } z_2(f) = \Sigma\, b(x, y) \cdot f(x, y)$$

$$\text{subject to } \sum_{y \in N} f(x, y) - \sum_{y \in N} f(y, x) = \begin{cases} 1 \text{ if } x = s, \\ -1 \text{ if } x = t, \\ 0 \text{ otherwise,} \end{cases}$$

(P2) $$f(x, y) \geqslant 0.$$

In the above formulation, it would be rare that the two objectives are simultaneously minimized. Therefore, the solution of the problem has to be based on the concept of nondominance [6]. Denote the feasible set of (P2) by F, and let $z(f)$ denote the two component vector $[z_1(f), z_2(f)]$. Let $f^* \in F$ and $G_{f^*} = f \in F: z(f) \leqslant z(f^*)$. The point f^* is said to be a *nondominated* solution if and only if $z(f^*) = z(f)$ for all $f \in G_{f^*}$. Extreme points of the set F which are nondominated solutions are called nondominated or efficient extreme points.

The following lemma provides the relationship between (P1) and (P2).

LEMMA 1: There exists a nondominated extreme point of (P2) which is a solution to (P1).

PROOF: Let f^* be that optimal solution to (P1) for which $\Sigma\, b(x, y) \cdot f(x, y)$ is minimized. Then, clearly, f^* is a nondominated solution to (P2). The unimodularity of the constraint matrix of (P2) implies that all the extreme points of F are integer vectors. Since each one of these extreme points of F is a vector of zeros and ones, it follows that any integer solution of (P2) must be an extreme point of F. Again, since f^* is an integer nondominated solution of (P2), it must be a nondominated extreme point of (P2). Thus an optimal solution to (P1) can be obtained by searching through the nondominated extreme points of the decision space F.

Let the feasible set in the two-dimensional objective space of (P2) be denoted by Z. Consider the following lemma:

LEMMA 2: The set Z is convex and each extreme point of Z corresponds to at least one extreme point of the feasible set F.

PROOF: The convexity of the set Z is obvious. Consider, now, an extreme point z° of the set Z. There exists at least one f, say f°, such that $z(f^\circ) = z^\circ$. Assume, contrary to the hypothesis, that there is no extreme point in the set F which corresponds to z°. Since f° is not an extreme point, we can write $f^\circ = \sum_{i=1}^{k} \alpha_i f_i$, $0 < \alpha_i < 1$, where the f_i's are distinct extreme points. So $z(f^\circ) = \sum_{i=1}^{k} \alpha_i z(f_i)$. If all $z(f_i)$'s are the same, then each must equal z°, leading to a contradiction. Otherwise, z°, an extreme point, is being expressed as a convex combination of some distinct points, which again leads to a contradiction.

Every extreme point of F need not correspond to an extreme point of Z. A point $\bar{z} \in Z$ is nondominated if and only if $z \leqslant \bar{z} \rightarrow z = \bar{z}$. Thus every nondominated point in the decision space F corresponds to a nondominated point in the objective space.

The algorithm presented below involves a search process in the set Z for nondominated extreme points. Each iteration involves the solution of a shortest chain problem in which the arc distance $c(x, y)$ is a positively weighted sum of $a(x, y)$ and $b(x, y)$. Initially, two nondominated extreme points are determined by solving two shortest chain problems, one for each objective. The positive weights are determined by the slope of the line joining these two points. If the two points are $z^{(r)}$ and $z^{(s)}$, the weights would be $w_1 = z_2^{(r)} - z_2^{(s)}$ and $w_2 = z_1^{(s)} - z_1^{(r)}$. With these newly defined costs on arcs, a shortest chain problem is solved. Thus, either a new nondominated point is revealed, or it is shown that the weighted objective function does not change. In the latter case, if $\bar{f}$ is a solution, then

$$\sum_{(x,y)\in A} c(x, y)\bar{f}(x, y) = w_1 \cdot z_1^{(r)} + w_2 \cdot z_2(s)$$

where $c(x, y) = w_1 \cdot z_1^{(r)} + w_2 \cdot z_2^{(s)}$. In this case, one of the alternative optima is the desired solution and the algorithm terminates. In the former case, the new nondominated point obtained is used to determine the positive weights for the next iteration.

3. ALGORITHM

STEP 0: Find $z_1^{(1)} = \text{Min}\,(z_1 | f \in F)$
and $z_2^{(1)} = \text{Min}\,(z_2 | z_1 = z_1^{(1)} \text{ and } f \in F)$.
[$z_2^{(1)}$ is obtained by searching through all the extreme points that yield $z_1^{(1)}$.]
Record $(z_1^{(1)}, z_2^{(1)})$. Similarly, find
$z_2^{(2)} = \text{Min}\,(z_2 | f \in F)$
and $z_1^{(2)} = \text{Min}\,(z_1 | z_2 = z_2^{(2)} \text{ and } f \in F)$.
Record $(z_1^{(2)}, z_2^{(2)})$ and set $k = 2$. Stop if

$B < z_2^{(2)}$ or $B \geqslant z_2^{(1)}$, since in the first case the problem is infeasible and in the second case the additional constraint is redundant. Otherwise, set $r = 1$ and $s = 2$ and go to Step 1.

STEP 1: Set $w_1^{(r,s)} = z_2^{(r)} - z_2^{(s)}$ and $w_2^{(r,s)} = z_1^{(s)} - z_1^{(r)}$.

Let $\bar{f}$ be the optimal solution to the shortest chain problem with $c(x, y) = w_1^{(r,s)} a(x, y) + w_2^{(r,s)} \cdot b(x, y)$ as the unit cost of arc (x, y). (If there are alternative optima, choose the one for which z_1 is minimum and call it $\bar{f}$.) If, for this solution $\bar{f}$, $\sum_{(x,y)\in A} c(x,y) \cdot \bar{f}(x, y) = w_1^{(r,s)} z_1^{(r)} + w_2^{(r,s)} z_2^{(r)}$, go to Step 3.

Otherwise set $k = k + 1$ and go to Step 2.

STEP 2: Let $z_1^{(k)} = \sum_{(x,y)\in A} a(x,y) \cdot \bar{f}(x, y)$

and $z_2^{(k)} = \sum_{(x,y)\in A} b(x,y) \cdot \bar{f}(x, y)$

If $z_2^{(k)} > B$, set $r = k$. If $z_2^{(k)} < B$, set $s = k$. If $z_2^{(k)} = B$, go to Step 4, otherwise go to Step 1.

STEP 3: Determine all the alternative optima for the shortest chain with $c(x, y)$, as defined in Step 1, as the unit cost on arc (x, y) and choose the one, say f^*, for which the additional constraint is satisfied and $\Sigma a(x, y) \cdot f^*(x, y)$ is minimized. Stop, since f^* is the desired solution.

STEP 4: Stop, since $\bar{f}$, the solution which yields $z_1^{(k)}$ and $z_2^{(k)}$ for the two objectives of (P2), is the desired solution.

Certain comments can be made about the algorithm. Both in Step 1 and Step 3, the algorithm would have to determine all the alternative optima to the shortest chain problem. This is not difficult to do, since all the integer solutions satisfying the constraints of (P2) are extreme points of (P2), and there are simplex-based algorithms for finding the shortest chain [1]. The simplex method can be used to obtain the shortest chain since it is easy to show the equivalence between the shortest chain and the assignment problem [5]. Also, Dijkstra's algorithm [2] can be modified to yield all the optimal solutions to the shortest chain problem.

4. VALIDITY OF THE ALGORITHM

Before proving the validity of the algorithm let us introduce some notation. Given two efficient extreme points $z^{(u)}, z^{(v)} \in R^2$, let $Q(u, v) = z | z \leqslant \lambda \cdot z^{(u)} + (1 - \lambda) z^{(v)}, z_1 \geqslant z_1^{(u)}, z_2 \geqslant z_2^{(v)}$. Let us assume that the algorithm does not converge in Step 0, so that problem (P1) has a nontrivial feasible solution.

LEMMA 3: Let z be a solution in the objective space corresponding to an optimal solution of (P1) which is a nondominated solution of (P2). Then, at each iteration of the algorithm, $z \in Q(r, s)$ with r and s as defined in the algorithm.

PROOF: Let us first show that the algorithm generates an efficient extreme point, $z^{(k)}$, at each iteration of the algorithm. Since the algorithm does not terminated at Step 0 of the first iteration, by assumption, $z^{(1)}$ and $z^{(2)}$ are distinct efficient extreme points. Thus w_1 and w_2 are strictly positive, and hence the new point, if any, obtained at Step 1 of the first iteration is the result of minimizing a positively weighted average of the two objective functions of (P2). Hence, this new point must be a nondominated extreme point. Thus, at each iteration $z^{(r)}$ and $z^{(s)}$ are efficient extreme points.

Since, at Step 0, $z_1^{(1)}$ minimizes $\Sigma a(x, y) \cdot f(x, y)$ and $z_2^{(2)}$, minimizes $\Sigma b(x, y) \cdot f(x, y)$, and $z^{(1)}$ and $z^{(2)}$ are nondominated extreme points, it follows that initially the result of Lemma 3 holds. Now assume that the result holds till iteration k. Then, as proved earlier, $z^{(k+1)}$ is a nondominated extreme point. Thus, the interior of the convex hull generated by $z^{(r)}$, $z^{(k+1)}$, and $z^{(s)}$ does not contain any efficient point. This is illustrated geometrically by Figure 1.

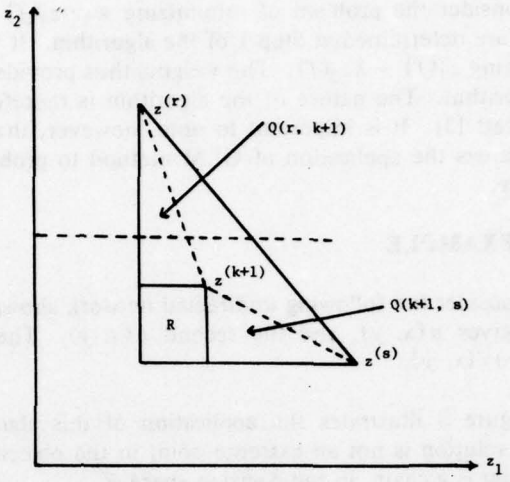


FIGURE 1 — Illustration of the proof.

Again, since rectangle R cannot contain any efficient points, $Q(r, k+1) \cup Q(k+1, s)$ contains the point z. However, depending upon the line $z_2 = B$, either $Q(r, k+1)$ or $Q(k+1, s)$ contains the desired point. Hence, $z \in Q(r, s)$ at every iteration of the algorithm.

The algorithm terminates when minimizing the objective function corresponding to the line joining $z^{(r)}$ and $z^{(s)}$ does not yield any new point. Thus $z|z = \{\lambda z^{(r)} + (1-\lambda)z^{(s)}\}$ defines a supporting hyperplane in the objective space Z. Also, from Lemma 3, $z \in Q(r, s)$ at termination. Thus $z = \lambda^0 z^{(r)} + (1-\lambda^0)z^{(s)}$, for some λ^0 between 0 and 1, provides an optimal solution to (P1). Step 3 of the algorithm determines a solution which corresponds to a point on the line joining $z^{(r)}$ and $z^{(s)}$ and which is closest to $z_2 = B$.

Figure 1 illustrates another important property of the algorithm. If the nondominated extreme points in the objective space are uniformly distributed over the entire nondominated region, each iteration of the algorithm divides the region of search by one half. Since 2^n is the maximal possible number of feasible solutions to (P1), it is a crude upper bound on the number of nondominated extreme points in the objective space. Hence, the maximum number of iterations needed to reduce the search to one efficient point would be $\log_2(2^n)(=n)$. Based upon Lemma 3 and the above remarks, we can state the following theorem.

THEOREM: The algorithm determines an optimal solution to (P1) in at most n iterations on the average.

5. RELATIONSHIP TO GENERALIZED LAGRANGE MULTIPLIER TECHNIQUE

$$\text{Defining} \qquad z_1(f) = \sum_{(x,y)\in A} a(x,y) \cdot f(x, y)$$

$$\text{and} \qquad z_2(f) = \sum_{(x,y)\in A} b(x,y) \cdot f(x, Y),$$

the problem (P1) can be rewritten as:

$$\begin{aligned} &\text{Minimize} && z_1(f) \\ &\text{subject to} && z_2(f) \leqslant B, \\ & && f \in F. \end{aligned}$$

Now consider the problem of minimizing $w_1 \cdot z_1(f) + w_2 \cdot z_2(f)$ subject to $f \in F$, where w_1 and w_2 are determined at Step 1 of the algorithm. If we define $\lambda = w_2/w_1$, this is equivalent to minimizing $z_1(f) + \lambda z_2(f)$. The weights thus provide a Lagrange multiplier at each iteration of the algorithm. The nature of the algorithm is therefore similar to the GLM method proposed by Everett [3]. It is important to note, however, that the algorithm presented here is a finite one whereas the application of GLM method to problem (P1) would not guarantee finite convergence.

6. AN EXAMPLE

Consider the following undirected network shown in Figure 2. The first number on an arc (x, y) gives $a(x, y)$, and the second $b(x, y)$. The additional constraint in the problem is $\Sigma b(x, y) f(x, y)$.

Figure 3 illustrates the application of this algorithm to the network in Figure 2. The desired solution is not an extreme point in the objective space, but corresponds to an extreme point, that is a chain, in the decision space F.

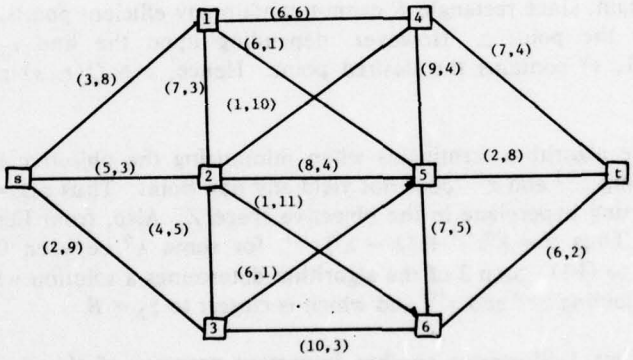


FIGURE 2 — A network.

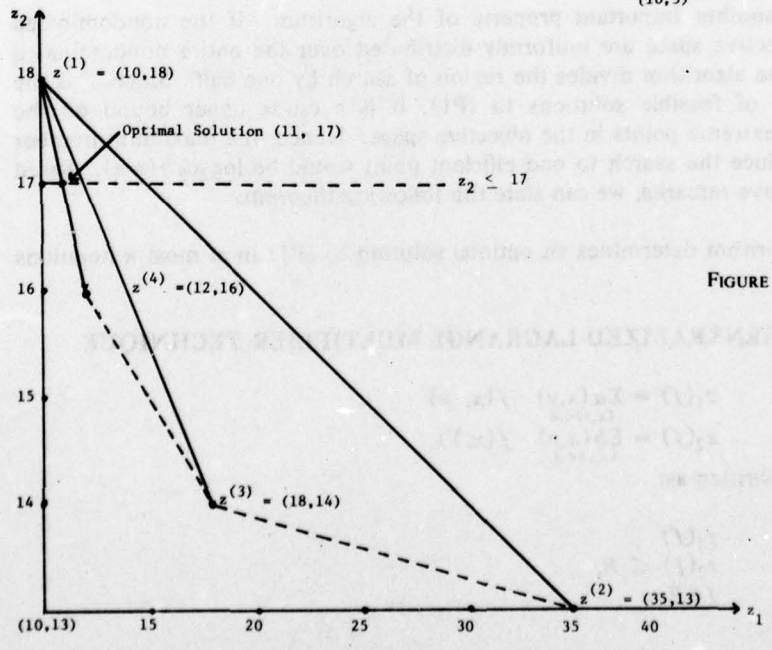


FIGURE 3 — Illustration of the algorithm

7. DISCUSSION

Although the constrained shortest chain problem is an integer problem, the algorithm presented in this paper would solve the problem efficiently without resorting to integer programming techniques. The methodology presented is quite general. An additional constraint can be handled easily in any 0-1 integer program. The algorithm becomes attractive when this integer program without the additional constraint has a special structure, such as unimodularity, leading to a computationally simple solution method. Minimum spanning tree, assignment, and knapsack problems would fall into this category.

The method can be easily extended to the situation where there is more than one additional constraint by "nesting." Thus, if there are two additional constraints, one would not have to solve more than n^2 shortest chain problems, on the average, to get the desired solution.

REFERENCES

[1] Dantzig, G.B., *Linear Programming and Extensions* (Princeton University Press, Princeton, N.J., 1962).

[2] Dijkstra, E.W., "A Note on Two Problems in Connection With Graphs," Numerische Mathematik *1*, 269-271 (1959).

[3] Everett, H., III, "Generalised Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," Operations Research, *11*, 399-417 (1963).

[4] Ford, L.R., Jr., and D.R. Fulkerson, *Flows in Networks* (Princeton University Press, Princeton, N.J., 1962).

[5] Wagner, H.M., *Principles of Operations Research* (Prentice Hall, Englewood Cliffs, N.J., 1969).

[6] Zeleny, M., *Linear Multiobjective Programming* (Springer-Verlag, New York, 1974).

PERMUTATION FLOW-SHOP THEORY REVISITED

Wlodzimierz Szwarc

School of Business Administration
University of Wisconsin-Milwaukee
Milwaukee, Wisconsin

ABSTRACT

The paper provides a new theoretical framework for generating dominance conditions and lower bounds and for solving special cases. All existing and new results have been derived in a routine and simple manner.

1. INTRODUCTION

Consider the following problem: Find a permutation $P = p_1,p_2,\ldots,p_n$ of rows $1,2,\ldots,n$ of a given $n \times m$ matrix $\{t_{rs}\}, t_{rs} > 0$ that minimizes

$$T(P,m) = \max_{w_1,w_2,\ldots,w_{n-1}} \left(\sum_{s=1}^{w_1} t_{p_1 s} + \sum_{s=w_1}^{w_2} t_{p_2 s} + \ldots + \sum_{s=w_{n-1}}^{m} t_{p_n s} \right) \tag{1}$$

over a set of integers $w_1,w_2,\ldots,w_{n-1}$ which satisfy conditions

$$1 \leqslant w_1 \leqslant w_2 \leqslant \ldots \leqslant w_{n-1} \leqslant m. \tag{2}$$

We have just formulated the well-known flow-shop problem of finding a permutation P that minimizes the completion time $T(P,m)$ of processing items $1,2\ldots n$ on machines $M_1,M_2,\ldots,M_m$ provided that the processing times t_{rs} of item r on machine M_s are given, that each item passes through the machines in the same order, $M_1,M_2,\ldots,M_m$, and that the machines process the items in the same order.

A sequence γ of cells of an $n \times m$ matrix is called a *segment* if each element (r,s) of γ (except the last) is followed by either $(r+1,s)$ or $(r,s+1)$. An $m+n-1$ element segment is called a *path*.

REMARK 1: Johnson [5] meant by a path a walk in the matrix from the upper left-hand corner to the lower right-hand corner, taking steps to the right or downward.

Observe that the summations in (1) are extended over the following path Γ (Γ makes a downward turn at each (r,w_r), $r < n$):

$$\Gamma = (1,1)\ldots(1,w_1);\ (2,w_1)\ldots(2,w_2);\ \ldots;\ (n,w_{n-1})\ldots(n,m).$$

$$T(P,m) = \max_{\Gamma \in [\Gamma]} \sum_{(r,s)\in\Gamma} t_{p_r s},$$

where $[\Gamma]$ is set of all paths. Path Γ is called a *critical path* for a given permutation P if $T(P,m) = \sum_{\Gamma} t_{p_r s}$.

This paper offers an entirely new theoretical framework for solving the problem that makes the shape of the segment its focal point. Then the derivation of all existing and new results in such areas as special cases [1], [3], [5], [8], [9], [11], [12], [14], [15], dominance conditions [2], [7], [10], [13], and lower bounds [4], [6] becomes a simple routine.

We have shown that the flow-shop problem can be formulated in pure combinatorial terms without the usual Machine-Job-Completion-Time interpretation. Hence, such widely used concepts as earliest time, shortest run-out time, and bottleneck and nonbottleneck machine are no longer necessary to develop lower bounds, since the shape of the critical path is the main factor. The author hopes that further extension of this new approach will ultimately produce a reasonably efficient branch-and-bound-solution method.

2. SELECTED TYPES OF CRITICAL PATHS

We find it convenient to describe a segment by a sequence of R and D symbols that indicate its right-hand and downward turns. Notice that consecutive symbols of this sequence are different. Figure 1 illustrates a DRDR segment.*

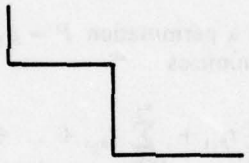


FIGURE 1

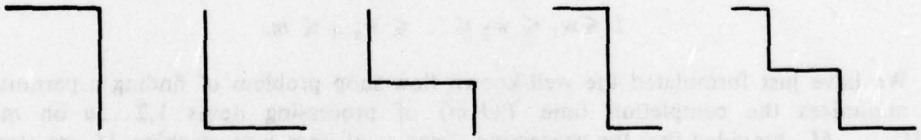


FIGURE 2

We will show that the solution of the flow-shop problem is easily available whenever the type of the critical path Γ remains the same for all permutations P and Γ is of the following type: (a) RD, (a') DR, (b) DRD, (b') RDR, (c) RDRD, (c') DRDR, or (d) RDRDR (see Figure 2).

REMARK 2: Each case includes D and R as a subcase. Cases a and a', b and b', c and c' are subcases of the subsequent cases (but not of each other).

CASE a: Path Γ passes row 1 and column m where $\sum_{r=1}^{n} t_{rm}$ is constant for each P.

SOLUTION: Find $\min_{r} \sum_{s=1}^{m-1} t_{rs} = \sum_{s=1}^{m-1} t_{1s}$. Then any $P = i \ldots,$† is the optimal permutation.

*Γ makes at most $m-1$ R-turns and $n-1$ D-turns.

†e.g., $P = i\pi$, where π is an arbitrary $n-1$ element permutation of numbers $r \in (1,2,\ldots,n) - (i)$.

CASE a': SOLUTION: Find $\min_r = \sum_{s=2}^{m} t_{rs} = \sum_{s=2}^{m} t_{is}$. Then any $P = , \ldots i^*$ is the optimal permutation.

CASE b: It is well-known that Johnson's algorithm solves the problem whenever Γ makes exactly one R turn (as in this case).

SOLUTION: Apply Johnson's method to the two-machine AB flow-shop problem with processing times $A_r = \sum_{s=1}^{m-1} t_{rs}$ and $B_r = \sum_{s=2}^{m} t_{rs}$.

REMARK 3: The solution of Case b remains optimal under a much relaxed assumption that Γ makes the single R turn for the permutation produced by Johnson's algorithm.

CASE b': Path Γ passes through an entire column, say column h.

SOLUTION: Find

$$\min_{p \neq q} \left[\sum_{s=1}^{h-1} t_{ps} + \sum_{s=h+1}^{m} t_{qs} \right] = \sum_{s=1}^{h-1} t_{is} + \sum_{s=h+1}^{m} t_{js}.$$

Then any $P = i \ldots j$ is the optimal permutation.

CASE d: Assume that Γ makes two D turns along columns u and v, $u < v$.

SOLUTION: Solve (as in Case b) the two-machine AB problem where $A_r = \sum_{s=u}^{v-1} t_{rs}$ and $B_r = \sum_{u+1}^{v} t_{rs}$, and let $1, 2, \ldots, n$ be the optimal permutation. Consider sequences $p \alpha q \overset{df}{=} p, 1, \ldots p-1, p+1, \ldots, q-1, q+1, \ldots, n, q$ for all possible $1 \leqslant p \neq q \leqslant n$. Let $t(p\alpha q)$ be the completion time of sequence $p\alpha q$ on machines A and B. Find

$$\min_{p \neq q} \left[\sum_{s=1}^{u-1} t_{ps} + \sum_{s=v+1}^{m} t_{qs} + t(p\alpha q) \right] = \sum_{s=1}^{u-1} t_{is} + \sum_{s=v+1}^{m} t_{js} + t(i\alpha j).$$

Then $P = i\alpha j$ is the optimal permutation.

CASES c and c': SOLUTION: Proceed as in Case d.

Sequence $P = i\beta \overset{df}{=} i, 1, \ldots, i-1, i+1, \ldots, n$ is optimal if

$$\min_{1 \leqslant p \leqslant n} \left[\sum_{s=1}^{u-1} t_{ps} + t(p\beta) \right] = \sum_{s=1}^{u-1} t_{is} + t(i\beta),$$

while sequence $P = \sigma j = 1, 2, \ldots, j-1, j+1, \ldots, n\, j$, is optimal if

$$\min_{1 \leqslant q \leqslant n} \left[\sum_{s=v+1}^{m} t_{qs} + t(\sigma q) \right] = \sum_{s=v+1}^{m} t_{js} + t(\sigma j).$$

*e.g., $P = \pi i$ (see previous footnote).

3. CLASSIFICATION OF SPECIAL CASES OF THE FLOW-SHOP PROBLEM

All known special cases belong to one of the following categories:

CASE I: For some $0 \leqslant h \leqslant m$,

$$\min_r t_{rs} \geqslant \max_r t_{rs+1}, \qquad 1 \leqslant s \leqslant h-1,$$

and

$$\min_r t_{rs+1} \geqslant \max_r t_{rs}, \qquad h+1 \leqslant s \leqslant m.$$

Several subcases of Case I have been solved in Refs. [3], [5], [9], [14], and [15].

CASE II [9]: For some $1 \leqslant h \leqslant m-1$,

$$\min_r t_{rs+1} \geqslant \max_r t_{rs}, \qquad 1 \leqslant s \leqslant h-1,$$

and

$$\min_r t_{rs} \geqslant \max_r t_{rs+1}, \qquad h+1 \leqslant s \leqslant m-1.$$

References [1], [3], [8], [14], and [15] considered special cases of II.

CASE III: For some $i \leqslant n$, and $h \leqslant m$,

$$t_{is} \geqslant t_{rs}, \qquad 1 \leqslant r \leqslant n,$$

and

$$t_{rh} \geqslant t_{rs}, \qquad 1 \leqslant s \leqslant m.$$

For subcases of Case III see Refs. [11], [12], and [15].

CASE IV: $\min(t_{r1}, t_{rm}) \geqslant t_{rs}$, $1 \leqslant r \leqslant n$, $2 \leqslant s \leqslant m-1$. Two subcases of Case IV have been solved in Ref. [15].

The envelope concept (see the following section) will be utilized in solving these special cases.

4. ENVELOPES

Let γ be an arbitrary segment having two endpoints (i,u) and (j, v), $i \leqslant j$, $u \leqslant v$. Such a segment consists of $j-i+v-u+1$ cells. Both RD and DR segments connecting the *same* endpoints as γ are called *envelopes* of γ and denoted by $\overline{\gamma}$ and $\underline{\gamma}$ respectively. Note each of the segments γ, $\overline{\gamma}$, and $\underline{\gamma}$ passes through $j-i+v-u+1$ cells. If $i=j$, or $u=v$, then $\gamma = \underline{\gamma} = \overline{\gamma}$.* Every path Γ can be presented as a sequence of segments $\gamma_1, \gamma_2, \ldots, \gamma_k$, where the last cell of γ_t is an initial cell of γ_{t+1}, $t = 1, \ldots, k-1$. Otherwise the γ_t are mutually exclusive. The envelope concept makes it easier to visualize the shape of Γ whenever it is presented as a sequence of segments. For instance, $\Gamma = \underline{\gamma}_1 \overline{\gamma}_2$ indicates that Γ is a DRD type (see Figure 3 where $\underline{\gamma}_1$ and $\overline{\gamma}_2$ share cell (r,s)), while $\Gamma = \overline{\Gamma}$ shows that Γ is a RD path.

*Then both envelopes are R or D segments, or single points (if $i=j$, and $u=v$).

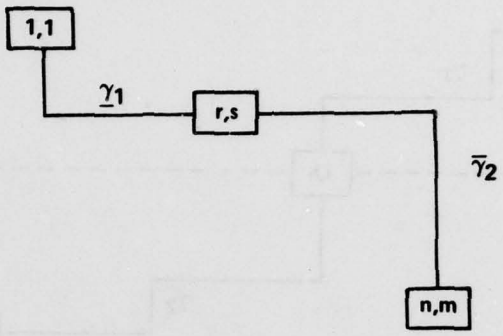


FIGURE 3

REMARK 4: Paths and segments will be denoted by symbols Γ and γ *exclusively.*

The solution of Cases I-IV requires a preliminary proof that, for every segment located in a certain area, one of the following conditions holds:

(3) $$\sum_{(r,s)\in\gamma} t_{p_r s} \leqslant \sum_{(r,s)\in\bar{\gamma}} t_{p_r s} \quad \text{for each } P,$$

or

(4) $$\sum_{\gamma} t_{p_r s} \leqslant \sum_{\underline{\gamma}} t_{p_r s} \quad \text{for each } P.$$

This proof always follows the same routine: We establish a one to one correspondence between cells $(r,s)\in\gamma$ and $(r',s')\in\bar{\gamma}$ (or $\underline{\gamma}$, depending on whether we want to prove (3) or (4)), and prove that, for each $(r,s)\in\gamma$,

(5) $$t_{p_r s} \leqslant t_{p_{r'} s'}.$$

We say that the *diagonal technique* or the row-column technique has been used if cells (r,s) and (r',s') are located on the same 45° diagonal ($r+s = r'+s'$), or on the same row ($r=r'$) or column ($s=s'$).

5. SOLUTION OF CASES I AND II

CASE I: We distinguish three subcases, $1 \leqslant h \leqslant m-1$, $h = 0$, and $h = m$.

1. Consider the first subcase: Let $\Gamma = \gamma_1\gamma_2$, where γ_1 and γ_2 share a cell $(i,h)\in\Gamma$ (see Figure 4). For $h=2$, $m-1$, this cell is specifically defined as: $(i,h) = \min_i(r,h)\in\Gamma$ (if $h=2$), and $(i,h) = \max_i(r,h)\in\Gamma$ (if $h = m-1$).* Define $\Gamma_0 = \underline{\gamma}_1\,\bar{\gamma}_2$. Applying the diagonal technique to prove (5), we can see that (3) and (4) hold for $\gamma = \gamma_2$ and $\gamma = \gamma_1$ respectively.

Hence,

$$\sum_{(r,s)\in\Gamma} t_{p_r s} \leqslant \sum_{(r,s)\in\Gamma_0} t_{p_r s}.$$

*Then segments γ_1 (if $h=2$) and γ_2 (if $h=m-1$) make a single R turn along row i.

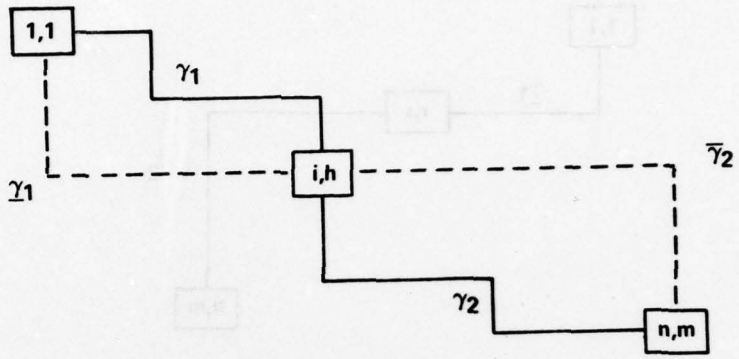


FIGURE 4

Consequently, the type of the critical path is DRD. For solution see Case b.

2. If $h = 0$, then $\min_r t_{rs+1} \geqslant \max_r t_{rs}$, $1 \leqslant s \leqslant m-1$.

Again applying the diagonal technique, we see that (3) holds for $\gamma = \Gamma$. Hence $\bar{\Gamma}$, a RD type, is a critical path. For solution see Case a.

3. If $h = m$, then $\min t_{rs} \geqslant \max_r t_{rs+1}$, $1 \leqslant s \leqslant m-1$.

Proceeding as in the previous case we can see that the critical path is a DR type (see Case a').

CASE II: Let $\gamma = \gamma_1\gamma_2\gamma_3$, where γ_1 and γ_2 share cell (i,h), while γ_2 and γ_3 share cell $(j,h+1)$. Consider $\Gamma_0 = \bar{\gamma}_1\ \gamma_2\ \underline{\gamma}_3$ (see Figure 5). Observe that γ_2 makes one R turn since it occupies two columns h and $h+1$ only. Utilizing the diagonal technique, we can see that (3) and (4) are satisfied for $\gamma = \gamma_1$ and $\gamma = \gamma_3$ respectively. Therefore

$$\sum_{\Gamma} t_{p_r s} \leqslant \sum_{\Gamma_0} t_{p_r s},$$

which means that RDRDR is the type of the critical path. For the solution, see Case d, where $u = h$, and $v = h+1$, $A_r = t_{rh}$ and $B_r = t_{rh+1}$.

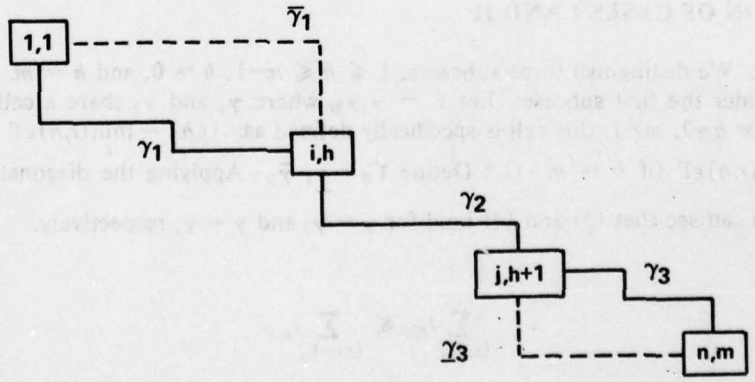


FIGURE 5

6. DECOMPOSITION OF THE FLOW-SHOP PROBLEM

CASE III: We will prove the following:

PROPERTY 1: Term $t_{ih} \overset{df}{=} \max_{r,s} t_{rs}$ appears in (1) for *every* P.

Proof: Let P = 1,2, ..., n. It is sufficient to show that cell (i,h) belongs to the critical path. Consider a path Γ that does not pass through (i,h) but crosses column h at (j,h) and row i at (i,u). Let $\Gamma = \gamma_1\gamma_2\gamma_3$ where γ_2 is a segment that connects end points (i,u) and (j,h). Define Γ_0 as $\gamma_1\gamma_2\gamma_3$ or $\gamma_1\bar{\gamma}_2\gamma_3$ depending on whether (i,h) is located below or above Γ. Obviously $(i,h)\epsilon\Gamma_0$. The row-column technique leads us to the conclusion that (4) (or (3)) holds for $\gamma = \gamma_2$. Thus

$$\sum_{\Gamma} t_{rs} \leqslant \sum_{\Gamma_0} t_{rs}, \qquad \text{Q.E.D.}$$

REMARK 5: It is easy to show (by a counterexample) that Case III is a necessary condition for Property 1.

Property 1 allows us to break the original problem into 2^{n-1} subproblems. This can be accomplished in the following manner: (1) Let $I = (1,2, \ldots, n)$, and generate all 2^{n-1} subsets J of $I-(i)$. (2) Let $\bar{J}$ denote a complement of J, and consider two flow-shop problems: one for jobset $J+(i)$ and machines $M_1M_2, \ldots, M_h$, the other for jobset $\bar{J}+(i)$ and machines $M_h, \ldots, M_m$. (3) Find sequences of the type αi and $i\beta$ that minimize the sum of processing times of both problems. Then permutation $P = \alpha i\beta$ is the optimal solution of the entire problem.

Consider Case III for $h = 1$ and $P = \ldots i$. The following obvious property holds:

PROPERTY 2: Sequence P maximizes (1).

PROOF: Observe that $\sum_{\Gamma} t_{p_r s} = \sum_{r=1}^{n} t_{r1} + \sum_{s=2}^{m} t_{is}$ is the largest value of (1) for all possible P.

7. ORDERED, SEMIORDERED AND OTHER SPECIAL CASES

ORDERED CASE [11]: (1) For each i and j, row i is either greater than, equal to, or smaller than row j (i.e., for all $1 \leqslant s \leqslant m$: either $t_{is} \geqslant t_{js}$, $t_{is}=t_{js}$, or $t_{is} \leqslant t_{js}$). (2) For any u and v, column u is either greater than, equal to, or smaller than column v (i.e., for all $1 \leqslant r \leqslant m$: either $t_{ru} \geqslant t_{rv}$, $t_{ru} = t_{rv}$, or $t_{ru} \leqslant t_{rv}$).

SEMIORDERED CASE [15]: The first condition of the Ordered Case is met. It is obvious that the Ordered Case is a subcase of Case III. Let h be the largest column. For $h = 1$ the following property holds [11]:

PROPERTY 3: (a) No critical path makes a R turn along row i if some subsequent row j (where P = $\ldots i \ldots j \ldots$) is greater than i.

(b) No critical path makes a D turn along column u if some subsequent column v is greater than u. Notice that Property 3b does not depend on permutation P.

Property 3 can be easily shown by applying the envelope approach along with the row-column technique from Section 3. According to Ref. [11], sequence P is optimal if all n rows are arranged in a nonincreasing order. Case $h = m$ is symmetrical to case $h = 1$. Hence the rows in the optimal permutation are arranged in a nondecreasing order. If $1 < h < m$, then the problem is decomposable as Case III (see Ref. [12]).

Let h=1. Assume also that m is the second largest column. Then Property 3 implies:

COROLLARY 1: The critical path makes a single R turn.

According to Ref. [15] (again with the help of the envelope approach) Corollary 1 is valid for the following less restricted cases: (a) Case IV, $m = 3$, and (b) Case IV, m arbitrary, provided this case is semiordered.

Consider the following two cases:

CASE 1: $t_{r1} = t_{r2} = \ldots = t_{rm}$ for each $1 \leqslant r \leqslant n$.

CASE 2: $t_{1s} = t_{2s} = \ldots = t_{ns}$ for each $1 \leqslant s \leqslant m$.

It is easy to see that the critical path for Case 1 makes, for each P, a single R turn along the largest row, say row i. Since $T(P,m) = \sum_{r=1}^{n} t_{r1} + \sum_{s=2}^{m} t_{is}$ is the same for each P, any sequence is optimal. The same conclusion is valid for Case 2, since (1) is constant for each permutation.

8. DOMINANCE CONDITIONS

Consider sequence $\rho = p_1 p_2, \ldots, p_k$, $\rho \subset I = (1,2,\ldots,n)$. Define

$$C(\rho,u,v) = \max_{u \leqslant w_1 \leqslant w_2 \leqslant \ldots w_{k-1} \leqslant v} \left[\sum_{s=u}^{w_1} t_{p_1 s} + \sum_{s=w_1}^{w_2} t_{p_2 s} + \ldots + \sum_{s=w_{k-1}}^{v} t_{p_k s} \right], \tag{6}$$

where $1 \leqslant u \leqslant v \leqslant m$. Then (see (1))

$$T(\rho,v) = C(\rho,1,v).$$

Divide ρ into t disjoint subsequences: $\alpha_1, \alpha_2, \ldots, \alpha_t \subset I$, $t \leqslant k$, preserving the order of the elements of ρ. The $\rho = \alpha_1 \alpha_2 \ldots \alpha_t$. From (1) and (6) we obtain the following formula:

$$T(\rho,v) = \max_{1 \leqslant u_1 \leqslant u_2 \leqslant \ldots \leqslant u_{t-1} \leqslant v} \left[T(\alpha_1,u_1) + C(\alpha_2,u_1,u_2) + \ldots + C(\alpha_t,u_{t-1},v) \right] \tag{7}$$

The following three dominance conditions assume that presequence $\sigma \subset I$ of permutation $P = \sigma\pi$ is fixed.

CONDITION I [4], [10]: $T(\sigma',u) \leqslant T(\sigma,u)$, $\forall\, 1 \leqslant u \leqslant m$, removes sequences $\sigma \ldots$ (sequences σ' and σ contain the same elements).

PROOF: Let π be an arbitrary sequence and $\pi \cap \sigma = \phi$.

In view of (7),

$$T(\sigma\pi,m) = \max_{1\leqslant u\leqslant m}[T(\sigma,u) + C(\pi,u,m)], \text{ and}$$

$$T(\sigma'\pi,m) = \max_{1\leqslant u\leqslant m}[T(\sigma',u) + C(\pi,u,m)].$$

Hence I $\Longrightarrow [T(\sigma'\pi,m) \leqslant T(\sigma\pi,m)]$, Q. E. D.

CONDITION II [3], [7], [13]: $T(\sigma ab,u) - T(\sigma b,u) \leqslant t_{av}$, $\forall\ 1\leqslant u\leqslant v\leqslant m$, where $a,b\epsilon I$, eliminates sequences $\sigma b\ldots$

PROOF: Consider two arbitrary disjoint sequences π' and π'' where

$\pi' \cup \pi'' = I - \sigma ab$. Again, by (7)

$$T(\sigma ab\pi'\pi'',m) = \max_{1\leqslant u\leqslant v\leqslant m}[T(\sigma ab,u) + C(\pi',u,v) + C(\pi'',v,m)], \text{ and}$$

$$T(\sigma b\pi' a\pi'',m) = \max_{1\leqslant u\leqslant v\leqslant w\leqslant m}[T(\sigma b,u) + C(\pi',u,v) + C(a,v,w) + C(\pi'',w,m)]$$

$$\geqslant \max_{1\leqslant u\leqslant v\leqslant m}[T(\sigma b,u) + C(\pi',u,v) + C(a,v,v) + C(\pi'',v,m)],$$

where $C(a,v,v) = t_{av}$ $\left[C(a,v,w) = \sum_{s=v}^{w} t_{as}\right]$.

Hence

(8) $$T(\sigma ab\pi'\pi'',m) \leqslant T(\sigma b\pi' a\pi'',m)$$

holds if for each u, and $v, u \leqslant v$. $T(\sigma ab,u) \leqslant T(\sigma b,u) + t_{av}$, which is Condition II.

CONDITION III [13]: $T(\sigma ab,u) - T(\sigma b,u) \leqslant t_{am}$, $\forall\ 1\leqslant u\leqslant m$, removes sequences $\sigma b\ldots a$.

PROOF: Assume $\pi'' = \phi$ in the preceding proof. Then it is obvious that (8) holds for $\pi'' = \phi$ whenever $T(\sigma ab,u) \leqslant T(\sigma b,u) + C(a,m,m) = T(\sigma b,u) + t_{am}$, $\forall\ u \leqslant m$, Q.E.D.

It is known (see Refs. [7] and [13]) that whenever Condition II is not met, there exists a *hypothetical* example where (8) is violated. To show it, assume

$$T(\sigma ab,u) > T(\sigma b,u) + t_{av} \qquad \text{for some } u \leqslant v,\ v \geqslant 2.$$

Define $\pi' = (p)$ and $\pi'' = (q)$. If p and q satisfy (8) and (9) of reference [13], that is:

$$t_{ps} \leqslant \begin{cases} \leqslant t_{bs+1}, & 1 \leqslant s \leqslant u-1, \\ \geqslant \max\,[T(\sigma ab,s+1) - T(\sigma ab,s),\ T(\sigma b,s+1) - T(\sigma b,s)], & u \leqslant s \leqslant v-1, \\ \geqslant T(\sigma bpa,s-1) - T(\sigma bp,s-1), & s = v, \\ > 0, & v+1 \leqslant s \leqslant m; \end{cases}$$

and

$$t_{qs} = \begin{cases} \leqslant \min\,(t_{as+1}, t_{ps+1}), & 1 \leqslant s \leqslant v-1, \\ \geqslant \max\,[T(\sigma abp,s+1) - T(\sigma abp,s),\ T(\sigma bpa,s+1) - T(\sigma bpa,s)], & v \leqslant s \leqslant m-1, \\ > 0, & s=m, \end{cases}$$

where the t_{ps} satisfy (8), then:

1. $T(\sigma abpq,m) = T(\sigma ab,u) + C(pq,u,m)$,

2. $C(pq,u,m) = \sum_{s=u}^{v} t_{ps} + \sum_{s=v}^{m} t_{qs}$,

3. $T(\sigma bpaq,m) = T(\sigma b,u) + C(paq,u,m)$, and

4. $C(paq,u,m) = \sum_{s=u}^{v} t_{ps} + t_{av} + \sum_{s=v}^{m} t_{qs}$.

Hence,

$$T(\sigma abpq,m) - T(\sigma bpaq,m) = T(\sigma ab,u) - T(\sigma b,u) - t_{av} > 0, \qquad \text{Q.E.D.}$$

One can generate the following dominance conditions that are independent of presequence σ, and involve elements $a,b \epsilon I$ only:

CONDITION IV [16]: $C(ab,u,v) \leqslant C(ba,u,v)$, $\forall\ 1 \leqslant u \leqslant v \leqslant m$, eliminates sequences $\ldots ba \ldots$.

PROOF: For any $\sigma,\pi \subset I$, $\sigma \cap \pi = \phi$,

$$T(\sigma ab\pi,m) = \max_{1 \leqslant u \leqslant v \leqslant m} [T(\sigma,u) + C(ab,u,v) + C(\pi,v,m)], \text{ and}$$

$$T(\sigma ba\pi,m) = \max_{1 \leqslant u \leqslant v \leqslant m} [T(\sigma,u) + C(ba,u,v) + C(\pi,v,m)].$$

Hence Condition IV $\Rightarrow [T(\sigma ab\pi,m) \leqslant T(\sigma ba\pi,m)]$, Q.E.D.

Condition IV can be verified in the following manner (see Ref. [16]): Consider a flow-shop problem where machines $M_1 M_2, \ldots, M_m$ pass through jobs a and b in order ab. If $M_1,M_2, \ldots, M_m$ is the optimal sequence obtained by Johnson's method, then Condition IV holds, and we eliminate sequences $\ldots ab \ldots$. If $M_m,M_{m-1}, \ldots, M_1$ is the optimal sequence, then we eliminate sequences $\ldots ba \ldots$ (Then Condition IV' holds — see Remark 6).

CONDITION V: $C(ab,k,u) - C(b,k,u) \leqslant t_{av}$, $\forall\ 1 \leqslant k \leqslant u \leqslant v \leqslant m$, eliminates sequences $\ldots b \ldots a \ldots$.

PROOF: Consider dominance condition II. Then

$$T(\sigma ab,u) - T(\sigma b,u) = \max_{1 \leqslant k \leqslant u} [T(\sigma,k) + C(ab,k,u)] - \max_{1 \leqslant k \leqslant u} [T(\sigma,k) + C(b,k,u,)]$$

$$\leqslant \max_{1 \leqslant k \leqslant u} [C(ab,k,u) - C(b,k,u)].$$

Hence V $\Longrightarrow T(\sigma ab,u) - T(\sigma b,u) \leqslant t_{av}]$, Q.E.D.

Condition V can be rewritten in the following form:

$$\max_{k \leqslant w \leqslant u} \left(\sum_{s=k}^{w} t_{as} - \sum_{s=k}^{w-1} t_{bs} \right) \leqslant t_{av},$$

which means that the total waiting time of item b for sequence ab processed on machines $M_k, \ldots, M_u$ does not exceed t_{av}.

REMARK 6: According to Ref. [13], pp. 1253-1254, one may arrange the machines and the items (i.e., the rows and columns of matrix $[t_{rs}]$) in a reversed order and find symmetrical dominance conditions.

Applying this approach to $\underline{\overline{V}}$ we get

CONDITION V′:

$$\max_{k \leqslant w \leqslant v} \left(\sum_{s=v}^{w} t_{bs} - \sum_{s=v}^{w+1} t_{as} \right) \leqslant t_{bu}, \quad V\, 1 \leqslant w \leqslant k \leqslant v \leqslant m,$$

which also eliminates sequences $\ldots b \ldots a. \ldots .$

Thus sequences $\ldots b \ldots a \ldots$ are removed whenever $\underline{\overline{V}}$ or $\underline{\overline{V}}'$ is met.

REMARK 7: Sequences $b \ldots$ can be eliminated if Conditions II ($\sigma = \phi$) or III′ hold.

All dominance conditions that involve a and b only (this includes II and III for $\sigma=\phi$) may improve the efficiency of the solution process, since they do not require excessive computations, and their verification may significantly reduce the number of branches.

We offer the following practical suggestions:

- Check the dominance conditions in the following order: (1) IV, IV′; (2) II, II′ ($\sigma = \phi$); (3) V, V′. Condition $\underline{\overline{V}}$ ($\underline{\overline{V}}'$) should not be checked if IV (IV′) or II (II′) does not hold.*

- The choice of a "forward" or "backward" procedure (See Remark 6) should depend on whether the number of sequences $b \ldots$ removed by Condition II exceeds the number of sequences $\ldots a$ removed by Condition II′.

- Prior to the branch-and-bound-solution procedure, find an initial solution P using Johnson's Method (as in Case b, Section 2).

Determine $T(P,m)$ and $t(P)$ the completion time of sequence P of the two-machine AB problem. If

$$T(P,m) = t(P) - \sum_{r=2}^{n-1} \sum_{s=2}^{m-1} t_{rs},$$

then P is an optimal solution (see Ref. [15]).

*Since Condition V is stronger than either IV or II while V′ is stronger than either IV′ or II′.

9. LOWER BOUNDS

Reference [6] generates lower bounds based on the bottleneck machine and non-bottleneck machine concept. Consider a permutation $P = \sigma\pi$. Without loss of generality, we may assume $\sigma = 1,2,\ldots,k$, $0 \leqslant k \leqslant n-2$. Let $[\pi]$ be a set of all $(n-k)!$ permutations of elements $k+1, k+2, \ldots, n$. In view of (6) and (7),

$$T(\sigma\pi,m) = \max_{1\leqslant w\leqslant m}[T(\sigma,w) + C(\pi,w,m)] \geqslant \max_{1\leqslant w\leqslant m}[T(\sigma,w) + \min_{\pi\in[\pi]} C(\pi,w,m)].$$

Define $[\gamma_w]$ as a set of all segments of the same type that connect end cells $(k+1, w)$ and (n,m). Observe that $[\gamma_w]$ is a single element segment if γ_w is a RD, DR, or D type (γ_m is always a D type).

Assume that the optimal sequence for the respective type (see Section 2) is $\pi_w = k+1, \ldots, n$. Then

$$\min_{\pi\in[\pi]} C(\pi,w,m) \geqslant \max_{\gamma_w\in[\gamma_w]} \sum_{(r,s)\in\gamma_w} t_{rs} \overset{df}{=} K_w.$$

Hence, the lower bound $LB(\sigma)$ for all possible permutations $\sigma \ldots$ is

$$LB(\sigma) = \max_{k+1\leqslant w\leqslant m}[T(\sigma,w) + K_w]. \tag{9}$$

If $\sigma = \phi$ $(k = 0)$ then (9) provides a lower bound for $T(P,m)$.

EXAMPLES:

1. Let γ_w is a RD segment. Then (see Case a)

$$K_w = \sum_{s=w}^{m-1} t_{is} + \sum_{r=k+1}^{n} t_{rm} = \min_{k+1\leqslant r\leqslant n}\left[\sum_{s=w}^{m-1} t_{rs} + \sum_{r=k+1}^{n} t_{rm}\right].$$

2. If γ sub w is a RDR segment then (see Case b′)

$$K_w = \max_{w\leqslant h\leqslant m}\left[\sum_{s=w}^{h-1} t_{is} + \sum_{r=k+1}^{n} t_{rh} + \sum_{s=h+1}^{m} t_{js}\right]$$

where

$$\sum_{s=w}^{h-1} t_{is} + \sum_{r=k+1}^{n} t_{rh} + \sum_{s=h+1}^{m} t_{js} = \min_{p\neq q}\left[\sum_{s=w}^{h-1} t_{ps} + \sum_{r=k+1}^{n} t_{rh} + \sum_{s=h+1} t_{qs}\right],$$

$k+1 \leqslant p, q \leqslant n$.

3. Assume that γ_w is a DRD type. According to Case b, the optimal solution π_w of the AB problem is $k+1, \ldots, n$.

$$K_w = \max_{k+1\leqslant q\leqslant n}\left[\sum_{r=k+1}^{q} t_{rw} + \sum_{s=w+1}^{m-1} t_{qs} + \sum_{r=q}^{n} t_{rm}\right].$$

Observe that $K_w = t(\pi_w) - \sum_{s=w+1}^{m-1} t_{rs}$ where $t(\pi_w)$ is the (minimum) processing time of sequence π_w of the two machine AB problem. Consequently,

$$LB(\sigma) = \max_{k+1\leqslant w\leqslant n}\left[T(\sigma,w) + t(\pi_w) - \sum_{s=w+1}^{m-1} t_{rs}\right].$$

Define $[\gamma_w]$ as a set of all segments of the same type that connect end cells $(k+1, w)$ and (n,m). Observe that $[\gamma_w]$ is a single element segment if γ_w is a RD, DR, or D type (γ_m is always a D type).

Assume that the optimal sequence for the respective type (see Section 2) is $\pi_w = k+1, \ldots, n$. Then

$$\min_{\pi\in[\pi]} C(\pi,w,m) \geqslant \max_{\gamma_w\in[\gamma_w]} \sum_{(r,s)\in\gamma_w} t_{rs} \stackrel{df}{=} K_w.$$

Hence, the lower bound LB(σ) for all possible permutations σ ... is

$$LB(\sigma) = \max_{k+\leqslant w\leqslant m} [T(\sigma,w) + K_w]. \tag{9}$$

If $\sigma = \phi$ $(k = 0)$ then (9) provides a lower bound for $T(P,m)$.

EXAMPLES:

1. Let γ_w is a RD segment. Then (see Case a)

$$K_w = \sum_{s=w}^{m-1} t_{is} + \sum_{r=k+1}^{n} t_{rm} = \min_{k+1\leqslant r\leqslant n} \left[\sum_{s=w}^{m-1} t_{rs} + \sum_{r=k+1}^{n} t_{rm}\right].$$

2. If γ sub w is a RDR segment then (see Case b')

$$K_w = \max_{w\leqslant h\leqslant m} \left[\sum_{s=w}^{h-1} t_{is} + \sum_{r=k+1}^{n} t_{rh} + \sum_{s=h+1}^{m} t_{js}\right]$$

where

$$\sum_{s=w}^{h-1} t_{is} + \sum_{r=k+1}^{n} t_{rh} + \sum_{s=h+1}^{m} t_{js} = \min_{p\neq q} \left[\sum_{s=w}^{h-1} t_{ps} + \sum_{r=k+1}^{n} t_{rh} + \sum_{s=h+1} t_{qs}\right],$$

$k+1 \leqslant p, q \leqslant n$.

3. Assume that γ_w is a DRD type. According to Case b, the optimal solution π_w of the AB problem is $k+1, \ldots, n$.

$$K_w = \max_{k+1\leqslant q\leqslant n} \left[\sum_{r=k+1}^{q} t_{rw} + \sum_{s=w+1}^{m-1} t_{qs} + \sum_{r=q}^{n} t_{rm}\right].$$

Observe that $K_w = t(\pi_w) - \sum_{r=k+1}^{n} \sum_{s=w+1}^{m-1} t_{rs}$ where $t(\pi_w)$ is the (minimum) processing time of sequence π_w of the two machine AB problem. Consequently,

$$LB(\sigma) = \max_{k+1\leqslant w\leqslant n} \left[T(\sigma,w) + t(\pi_w) - \sum_{r=k+1}^{n} \sum_{s=w+1}^{m-1} t_{rs}\right].$$

Define LB_x as $LB(\sigma)$ where x indicates the type of γ_w. In view of Remark 2 and the definition of K_w, the following inequalities hold for any σ:

$$\left\{\begin{array}{l} \max(LB_a, LB_{a'}) \leqslant \min(LB_b, LB_{b'}), \\ \max(LB_b, LB_{b'}) \leqslant \min(LB_c, LB_{c'}), \text{ and} \\ \max(LB_c, LB_{c'}) \leqslant LB_d. \end{array}\right\} \tag{10}$$

REFERENCES

[1] Arthanari, T.S., and A.C. Mukhopadhyay, "A Note on a Paper by W. Szwarc," Naval Research Logistics Quarterly **15**, 135-138 (1971).

[2] Gupta, J.N.D., "An Improved Combinatorial Algorithm for the Flowshop Scheduling Problem, Operations Research, **19**, 1753-1758 (1971).

[3] Gupta, J.N.D., "Optimal Schedules for Special Structure Flowshops," Naval Research Logistics Quarterly **22**, 255-269 (1975).

[4] Ignall, E., and L. Schrage, "Application of the Branch and Bound Technique to Some Flow-Shop Scheduling Problems" Operations Research, **13**, 400-412 (1965).

[5] Johnson, S.M., "Optimal Two and Three-Stage Production Schedules with Setup Times Included," Naval Research Logistics Quarterly **1**, 61-68 (1954).

[6] Lenstra, J.K., *Sequencing By Enumerative Methods*, Chapter 12, (Mathematisch Centrum, Amsterdam, 1976).

[7] McMahon, G.B., "Optimal Production Schedules for Flow Shops," Canadian Operational Society Journal, **7**, 141-151 (1969).

[8] Nabeshima, I., "The Order of n Items Processed on m Machines," Journal of Operations Research Society of Japan, **3**, 170-175 (1960), and **4**, 1-8 (1961).

[9] Nabeshima, I., "Notes on the Analytical Results in Flow Shop Scheduling Problem," Parts 1 and 2, Reports of the University of Electro-Communications, **27**, 245-252 and 253-257 (1977).

[10] Smith, R.D., and R.A. Dudek, "A General Algorithm for Solution of the n-Job M-Machine Sequencing Problem of the Flow Shop," Operations Research, **15**, 71-82 (1967).

[11] Smith, M.L., S.S. Panwalkar, and R.A. Dudek, "Flow Shop Sequencing with Ordered Processing Time Martrices," Management Science, **21**, 544-549 (1975).

[12] Smith, M.L., S.S. Panwalkar, and R.A. Dudek, "Flowshop Sequencing Problem with Ordered Processing Time Matrices: A General Case," Naval Research Logistics Quarterly, **23**, 481-486 (1976).

[13] Szwarc, W., "Optimal Elimination Methods in the $m \times n$ Flow-Shop Scheduling Problem," Operations Research, **21**, 1250-1259 (1971).

[14] Szwarc, W., "Mathematical Aspects of the $3 \times n$ Job-Shop Sequencing Problem," Naval Research Logistics Quarterly, **21**, 145-153 (1974).

[15] Szwarc, W., "Special Cases of the Flow-Shop Problem," Naval Research Logistics Quarterly, **24**, 483-492, (1977).

[16] Szwarc, W., "Precedence Relations of the Flow-Shop Problem," submitted to Operations Research.

AN ALGORITHM FOR 0-1 MULTIPLE-KNAPSACK PROBLEMS

Ming S. Hung

Cleveland State University
Cleveland, Ohio

John C. Fisk

State Universty of New York at Albany
Albany, New York

ABSTRACT

The 0-1 multiple-knapsack problem is an extension of the well-known 0-1 knapsack problem. It is a problem of assigning m objects, each having a value and a weight, to n knapsacks in such a way that the total weight in each knapsack is less than its capacity limit and the total value in the knapsacks is maximized.

A branch-and-bound algorithm for solving the problem is developed and tested. Branching rules that avoid the search of redundant partial solutions are used in the algorithm. Various bounding techniques, including Lagrangean and surrogate relaxations, are investigated and compared.

1. INTRODUCTION

The 0-1 knapsack problem is a familiar one in operations research. In its simplest form, the problem is to find the most desirable set of objects to place in a single container of given capacity. Eilon and Christofides [4] describe a series of generalizations to the 0-1 knapsack problem, each of which they define as a type of 0-1 loading problem. The type of 0-1 loading problem we address in this paper may be formulated as follows:

$$\text{(1)} \quad \text{(P) maximize } z = \sum_i \sum_j c_i x_{ij}$$

$$\text{(2)} \quad \text{subject } \sum_i w_i x_{ij} \leqslant b_j \text{ for all } j,$$

$$\text{(3)} \quad \sum_j x_{ij} \leqslant 1 \text{ for all } i,$$

$$\text{(4)} \quad x_{ij} = 0, 1 \text{ for all } i,j,$$

where c_i and w_i represent the value and weight, respectively, of each of m objects and b_j represents the capacity constraint on each of n containers. Our objective is to choose objects for placement into containers so as to maximize the total value of the objects chosen without violating any of the capacity constraints b_j. We assume that the objects are indexed such that $c_1/w_1 \geqslant c_2/w_2 \geqslant \ldots \geqslant c_m/w_m$. We define this particular type of 0-1 loading problem as the 0-1 multiple-knapsack problem.

A great deal of research has been done on various forms and generalizations of the 0-1 knapsack problem (see Salkin and de Kluyver [18] for a survey of the knapsack problem and related problems). Little work has been done on (P), however. Eilon and Christofides discuss the application of problem forms such as (P) but do not suggest a solution. In a more recent paper, Ingargiola and Korsh [15] describe a series of rules useful for limiting the search for optimality, though little computational experience is given to indicate the effectiveness of these rules.

In this paper, we will present a branch-and-bound procedure for solving (P). Emphasis is on an empirical computational study of various new bounding techniques, including Lagrangean relaxation and surrogate relaxation. Section 2 discusses the relaxations and the solution of their multipliers. Section 3 presents the algorithm in detail. After a general description, two subsections discuss more difficult parts of the algorithm. Section 4 presents computational results that indicate the relative efficiency of each of the relaxation procedures presented in Section 2.

2. RELAXATIONS OF THE MULTIPLE-KNAPSACK PROBLEM

In this section we discuss two relaxation methods found to be useful in providing tight bounds for discrete optimization problems, the Lagrangean and the surrogate relaxations.

We define the Lagrangean relaxation of (P) relative to a nonnegative vector $\lambda=(\lambda_i)$ to be

$$(5)\quad (PR_\lambda)\quad \text{maximize } \sum_i \sum_j c_i x_{ij} - \sum_i \lambda_i \Big(\sum_j x_{ij} - 1\Big)$$

$$= \sum_i (c_i - \lambda_i) \sum_j x_{ij} + \sum_i \lambda_i$$

$$(6)\quad \text{subject to } \sum_i w_i x_{ij} \leqslant b_j \text{ for all } j,$$

$$(7)\quad x_{ij} = 0,1 \text{ for all } i,j.$$

An important feature of the relaxed problem (PR_λ) is that, once values of $\lambda = (\lambda_i)$ are found, it decomposes into a collection of single-knapsack problems, one for each j. Ross and Soland [17] used a similar approach to solve the generalized problem.

Another potentially useful relaxation of (P) is surrogate relaxation. Unlike the Lagrangean solution strategy, which absorbs a set of constraints into the objective function, this strategy replaces the original set of constraints by a new one called a surrogate constraint. The concept and the applicability of the surrogate procedure were introduced by Glover [10], [11], and useful refinements of the procedure were suggested by Balas [2], and Geoffrion [6]. Additional work in this area has been presented by Greenberg [12] and Greenberg and Pierskalla [13], among others.

The surrogate relaxation of (P) can be defined as

$$(8)\quad (PR_\pi)\quad \text{maximize } \sum_i \sum_j c_i x_{ij}$$

$$(9)\quad \text{subject to } \sum_j \sum_i \pi_j w_i x_{ij} \leqslant \sum_j \pi_j b_j.$$

$$(10)\quad \sum_j x_{ij} \leqslant 1, \text{ for all } i$$

$$(11)\quad x_{ij} = 0,1 \text{ for all } i,j.$$

Surrogate relaxation converts problem (P) into a single-knapsack problem. This can be observed if we replace $\sum_j x_{ij}$ with y_i in (8) through (10), thus reducing (PR_π) to

$$(12)\quad (PR_\pi') \text{ maximize } \sum_i c_i y_i$$

$$(13)\quad \text{subject to } \sum_j \sum_i \pi_j w_i y_i \leqslant \sum_j \pi_j b_j,$$

$$(14)\quad y_i = 0, 1 \text{ for all } i.$$

It has been pointed out (Geoffrion [7], Greenberg and Pierskalla [13]) that such relaxations can provide a bound closer to the optimal value of the integer solution than does the linear programming relaxation. The tightness of the bound clearly depends on the choice of the multipliers; i.e., λ for (PR_λ) and π for (PR_π). One suitable choice is to set them equal to the optimal dual multipliers of the continuous (linear programming) problem of (P). We see in the following paragraphs that these dual multipliers are easily found.

Let $(\bar{P})$ denote the continuous relaxation of (P) by replacing (4) with $1 \geqslant x_{ij} \geqslant 0$ for all i, j. An optimal solution to $(\bar{P})$ can be readily found because $(\bar{P})$ contains the properties that Dantzig [3] identified in the 0-1 single-knapsack problem. Let $(\bar{x}_{ij})$ denote optimal solution to $(\bar{P})$. Since the objects are ordered such that $c_1/w_1 \geqslant c_2/w_2 > \ldots \geqslant c_m/w_m$, we obtain the following results for $(\bar{x}_{ij})$:

$$(16)\quad \sum_j \bar{x}_{ij} = \begin{cases} 1 & \text{if } i < t \\ \left(\sum_j b_j - \sum_{i \leqslant t-1} w_i\right) \Big/ w_t & \text{if } i = t \\ 0 & \text{if } i > t \end{cases}$$

where t is the smallest object index such that

$$(17)\quad \sum_{i \leqslant t} w_i > \sum_j b_j.$$

Let $(\bar{\lambda}_i)$ and $(\bar{\pi}_j)$ be, respectively, the optimal dual multipliers of constraints (2) and (3) in $(\bar{P})$. By the complementary slackness theorem

$$(18)\quad \bar{\lambda}_i = \begin{cases} c_i - w_i(c_t/w_t) & \text{if } i < t, \\ 0 & \text{if } i \geqslant t, \end{cases}$$

and

$$(19)\quad \bar{\pi}_j = c_t/w_t \text{ for all } j.$$

The term $(PR_{\bar{\lambda}})$ denotes the Lagrangean problem when $\bar{\lambda} = (\bar{\lambda}_i)$ is used as the multiplier vector. As noted before, $(PR_{\bar{\lambda}})$ decomposes into a series of single-knapsack problems, one for each j. Similarly, when $(\bar{\pi}_j)$ are used in the surrogate problem (PR_π) to obtain $(PR_{\bar{\pi}}')$, a single-knapsack problem is defined as

$$(20)\quad (PR_{\bar{\pi}}) \max \quad \sum_i c_i y_i$$

$$(21)\quad \text{subject to } \sum_i w_i y_i \leqslant \sum_j b_j,$$

$$(22)\quad y_i = 0, 1 \text{ for all } i.$$

Although other multipliers may provide better bounds than $\bar{\lambda}$ and $\bar{\pi}$, we feel that, since $\bar{\lambda}$ and $\bar{\pi}$ are so easy to compute, the extra computional effort required to find other multipliers would overshadow the savings realized. It is important to note that the bounds derived from $(\mathrm{PR}_{\bar{\lambda}})$ or $(\mathrm{PR}_{\bar{\pi}})$ are always better than or equal to the bounds from linear programming relaxation.

3. The Branch-and-Bound Algorithm

The essential steps of our algorithm are closely related to those developed for the 0-1 single-knapsack problem by Ahrens and Finke [1] among others [5,16]. The general form of these algorithms can best be described as being of a "branch and exclude" type which successively determines whether or not an object is to be in the knapsack. The presence of multiple knapsacks in our problem demands a further decision: To what knapsack is the object to be assigned next?

In this paper, we construct successively higher levels of the branch-and-bound tree either by assigning an object to a knapsack or by excluding that object from all knapsacks. Any chosen level of our branch-and-bound tree, therefore, represents the assignment of some objects to the knapsacks and the exclusion of some other objects from all knapsacks. For simplicity, assume that the excluded objects have been assigned to the $(n+1)^{st}$ (dummy) knapsack. Let F denote the index set of the objects that have not been assigned. When $F=\phi$, we have found a feasible solution and the corresponding objective value z, which can then be compared to the incumbent solution value z^*.

Define S to be the index set of the objects that have been assigned: $S = \left\{ i \middle| \sum_{j=1}^{n+1} x_{ij} = 1 \right\}$. If M is the index set of all objects, then $S \cup F = M$ and $S \cap F = \phi$. The cardinality of set S thus coincides with the level of the branching tree. To avoid excessive notation, we simply note that if $S \neq \phi$, then (P) and its relaxed problems (PR_{λ}) and (PR_{π}) are reduced to problems with respect to objects in F only.

A general description of the steps of our algorithm is as follows:

STEP 1 (Initialization): Set $z^* = -\infty$, $S = 0$, $F = M$. Let tree-level index $k = 1$.

STEP 2 (Bounding): Solve the relaxed problem $(\mathrm{PR}_{\bar{\lambda}})$ or $(\mathrm{PR}_{\bar{\pi}})$ with respect to F. Compute the value of such solution, $\bar{z}_k$, by utilizing the original objective function (1). If $\bar{z}_k \leqslant z^*$, go to Step 5. If the solution is feasible to (P), go to Step 4; otherwise, go to Step 3.

STEP 3 (Branching): Select an object i and assign the object to one of the knapsacks (including the dummy knapsack). Record the value of the objects in the knapsacks (excluding the dummy). If all objects have been assigned, go to Step 4, otherwise update S,F, and k and go to Step 3.

STEP 4: Update z^* and its corresponding solution. Go to Step 5.

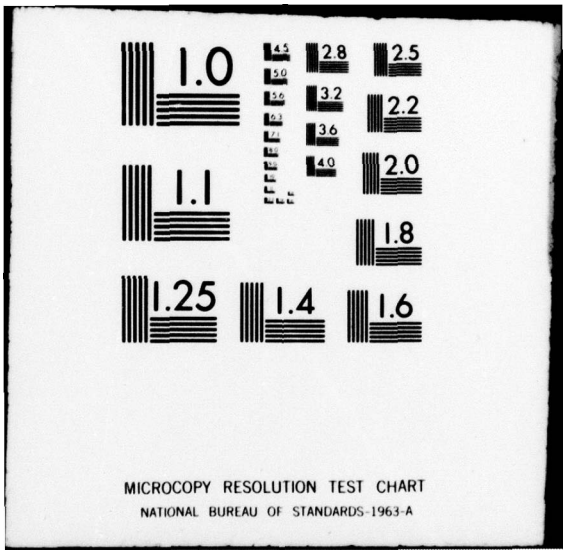
STEP 5: (Backtracking): Find the smallest level k such that $\bar{z}_{k_0} \leqslant z^*$. Denote the level preceding k_0 as k_{-1} (i.e., $k_{-1} = k_0 - 1$), and denote the corresponding object index in S as $i_{k_{-1}}$. If $k_{-1} \leqslant 0$, terminate the procedure, otherwise set $k = k_{-1}$ and free all indices in S following $i_{k_{-1}}$. Finally, assign object $i_{k_{-1}}$ to a knapsack that is different from the knapsacks to which it has

AD-A064 991 OFFICE OF NAVAL RESEARCH ARLINGTON VA F/G 15/5
NAVAL RESEARCH LOGISTICS QUARTERLY. VOLUME 25, NUMBER 3.(U)
SEP 78

UNCLASSIFIED NL

3 OF 3
AD
A064991

END
DATE
FILMED
4 -79
DDC


1.0
2.8
2.5
3.2
2.2
3.6
2.0
4.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

previously been assigned and go to Step 2. If object i_{k-1} has been previously assigned to each of the $n+1$ boxes, however, set $\bar{z}_{k-1} = -\infty$ and go to Step 5.

Detailed descriptions of steps 2,3, and 5 above are provided in the following two subsections.

3a: Solving the relaxed problems (Step 2)

Both the Lagrangean relaxation and the surrogate relaxation reduce the original problem (P) to single-knapsack problems. Our single-knapsack algorithm is adopted from program β of Ahrens and Finke [1]. The algorithm is easy to use and has performed satisfactorily.

A few rules were added to the single-knapsack algorithm to increase efficiency. At a given level of our branching tree, some of the knapsacks have been assigned objects. For each knapsack j, let f_j be its unassigned capacity, that is

$$f_j = b_j - \sum_{i \in S} w_i x_{ij} \text{ for all } j. \tag{23}$$

RULE 1: Knapsack j is deleted from (PR_λ) and (PR_π) if $f_j < \min_{i \in F} w_i$, $j = 1, \ldots, n$.

RULE 2: Object i is deleted from (PR_π) if $w_i > \max_j f_j, i \in F$

RULE 3: For each single knapsack j in (PR_λ), object i is deleted from consideration if $w_i > f_j$, $i \in F$.

We have found such simple rules to be very effective in reducing the amount of computation.

3b. Assigning objects to knapsacks (Steps 3 and 5)

In considering Steps 3 and 5, we must make two important decisions: (1) which object to choose for branching (Step 3) and (2) to which knapsack should an object be assigned (Steps 3 and 5).

The branching object we choose depends upon the type of relaxation employed. In the case of Lagrangean relaxation, we choose that object in F which appears in the most single-knapsack solutions involved in calculating $(PR_{\bar{\lambda}})$. In the event of ties, the lowest-indexed object in F is chosen. This choice of branching object has two advantages. First, it has the ability to change the most-violated constraint in (3) of the original problem into a satisfied constraint. Second, it will generally generate a tighter upper bound for use at the next level of the branching tree.

Since the surrogate relaxation does not solve a series of single-knapsack problems to calculate a bound, the object in F which we choose to branch on is simply the lowest-indexed one. The object chosen by this rule represents the object that has the highest value-to-weight ratio among objects not yet assigned and thus has the highest a priori chance of being in the optimal solution of (P).

When considering the assignment of a chosen object to a knapsack, we use several rules developed by Hung and Brown [14] for this type of assignment. The rules are useful to reduce the number of knapsacks an object may be assigned to.

Assume that objects of the same weight belong to a class and that knapsacks of the same capacity belong to a class. The order of objects in a class is determined by the order in which they are chosen as branching objects. (This ordering does not change the indexing determined by value-to-weight ratios at the beginning of the algorithm.) However, we assume that knapsacks are indexed in descending order of their capacity, i.e., $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n$. Therefore, knapsacks of the same capacity class are grouped together. For simplicity of expression, by a "preceding object" to object i we mean an object in the same class as object i whose order in the class precedes object i.

The rules for assigning an object i to knapsacks are

RULE 4: Object i can be assigned to knapsack j if either (a) object i is the first object in its class to be considered or (b) the preceding object was assigned to a knapsack whose index is not greater than j.

RULE 5: Object i can be assigned to knapsack j if either (a) j is the smallest index in its class or (b) knapsack $j-1$ is not empty.

RULE 6: Object i can be assigned to knapsack j if $w_i \leqslant f_j$, where f_j is the remaining capacity of knapsack j as defined in (21).

For ease of bookkeeping and backtracking, we define μ_i to be an index set of all knapsacks to which object i can be assigned. Note that if an object cannot be assigned to any knapsack j, $1 \leqslant j \leqslant n$, according to Rules 4 through 6 we let μ_i contain the index of the dummy knapsack only. Once μ_i is determined, we assign object i to the knapsack which has the smallest index in μ_i and then remove the index from μ_i. When we return to object i during the backtracking step (Step 5), we simply reassign object i to the lowest-indexed knapsack remaining in μ_i.

For a detailed discussion of the branching rules 4 through 6, the reader is referred to Hung and Brown [14].

4. COMPUTATIONAL EXPERIENCE

The algorithm as described has been programmed in FORTRAN V code and run on a UNIVAC 1108. We obtained series of problems consisting of up to 200 objects and up to six boxes by generating values and weights independently from a uniform distribution in the interval [10,100]. Box capacities were then generated in a similar manner, except that the interval $b_l \leqslant b_j \leqslant b_u$ was used, where $b_l = [0.4(\Sigma w_i/n)]$ and $b_u = [0.6(\Sigma w_i/n)]$*. The final box capacity generated, b_n, was chosen such that occupancy ratio $= \Sigma b_j / \Sigma w_i = 0.5$. If $b_j < \min_i w_i$, or $\max_j b_j < \max_i w_i$, the set of generated box capacities was discarded and a new set was generated. The occupancy ratio of 0.5 was used for all problems attempted.

Tables 1 and 2 indicate computation times for our algorithm, with first the surrogate relaxation (Table 1) then the Lagrangean relaxation (Table 2), to determine upper bounds on

*The brackets [ ] denote the greatest integer less than the enclosed quantity.

TABLE 1 — *Surrogate Relaxation*

Number of Objects	Two Boxes				Three Boxes				Four Boxes			
	Complete Solutions	Solution Time (S)			Complete Solutions	Solution Time (S)			Complete Solutions	Solution Time (S)		
		High	Low	Av.		High	Low	Av.		High	Low	Av.
20	10	0.2	0.01	0.1	10	4.1	0.01	0.4	10	33.6	0.3	9.6
30	10	0.20	0.01	0.1	10	5.4	0.03	0.8	10	75.9	0.3	22.3
40	10	0.6	0.04	0.2	10	52.5	0.4	19.6	1	–	–	–
60	10	2.0	0.1	0.7	1	–	–	–	1	–	–	–
80	3	–	–	–	0	–	–	–	0	–	–	–
100	0	–	–	–	0	–	–	–	0	–	–	–

TABLE 2 — *Lagrangean Relaxation*

Number of Objects	Two Boxes				Three Boxes				Four Boxes			
	Complete Solutions	Solution Time (S)			Complete Solutions	Solution Time (S)			Complete Solutions	Solution Time (S)		
		High	Low	Av.		High	Low	Av.		High	Low	Av.
20	10	1.5	0.1	0.4	10	35.1	0.2	4.0	10	90.1	0.8	18.6
30	10	5.6	0.3	1.1	10	3.7	0.3	0.9	10	83.3	0.4	10.9
40	10	5.1	0.8	2.3	10	80.2	0.7	16.1	4	–	–	–
60	10	7.4	1.0	2.8	10	38.1	1.7	6.8	7	–	–	–
80	10	21.0	2.2	7.2	10	14.2	2.7	6.6	5	–	–	–
100	10	11.0	3.9	6.5	10	9.7	4.3	7.1	4	–	–	–

partial solutions. The sets of rules defined in Sections 3a and 3b were used in each case. For each object/box combination a total of ten problems was attempted, and the total number of complete solutions obtained within a maximum of 250-s running time, excluding input/output time, is specified. For those problem sets requiring $\leqslant$ 250-s for completion of ten problems, the maximum, minimum, and average solution times are also specified. Due to the method in which problem sets were generated, each relaxation procedure attempted to solve exactly the same set of problems.

The results summarized in Tables 1 and 2 indicate that the surrogate relaxation is relatively more efficient when the number of objects and/or boxes is small, but that the Lagrangean relaxation is vastly superior when one attempts to solve larger problems. Also, our experience, as shown in Tables 1 and 2, would indicate that, although the number of objects is an important factor in estimating solution time, the number of boxes has a tremendous influence*. A further indication of the effect of the number of boxes can be shown by the results obtained in the attempt to solve an additional set of ten problems consisting of 20 objects and six boxes (occupancy ratio = 0.5). The computer ran out of time after having solved only 3 of these problems.

In an attempt to improve upon the solution times presented in Tables 1 and 2, we solved a series of problems in which we applied both the surrogate relaxation and the Lagrangean relaxation for each candidate problem then used the bound having minimum value as a test against the incumbent. While the number of branches in the branch-and-bound tree was in general reduced with this method, the results were not significantly better and are not reported here.

*This should not be surprising, since the number of distinct solutions to (P) can be characterized by $(n+1)^m$. For a 20-object problem, then, the maximum number of solutions, if we use only one box, is 1.04×10^6, but for six boxes this number increases to more than 7.9×10^{16}.

As has been shown in Tables 1 and 2, it becomes increasingly difficult to obtain optimal answers to (P) when the problem size becomes large. We were, therefore, curious to see how efficient our Lagrangean relaxation routine would be if we were to require solutions which were within some small percentage value α of a known upper bound on the optimal solution to (P). Since each of the initial surrogate and Lagrangean relaxations yields such an upper bound, we determined to specify $\alpha = 0.5\%$ and to run a series of problems generated in the usual manner but to suspend calculations when we found a solution value $z \geqslant R(1-\alpha/100)$, where R represents the minimum value associated with the initial surrogate and Lagrangean relaxations. The results of these calculations are shown in Table 3.

As can be seen from Table 3, our Lagrangean relaxation technique becomes very efficient when we require only that our solutions lie within some small value of the known upper bound. These results are very encouraging when they are considered in the light of the obvious difficulty in obtaining provably optimal solutions to large problems.

TABLE 3 — *Lagrangean Relaxation* ($\alpha = 0.5\%$)

Number of Objects	Two Boxes				Four Boxes				Six Boxes			
	Complete Solutions	Solution Time (S)			Complete Solutions	Solution Time (S)			Complete Solutions	Solution Time (S)		
		High	Low	Av.		High	Low	Av.		High	Low	Av.
50	10	2.3	1.3	1.6	10	3.3	1.3	2.0	10	98.6	1.4	15.4
100	10	4.5	3.1	4.4	10	9.1	4.4	5.8	10	13.3	5.6	7.3
150	10	14.4	8.8	11.0	10	16.3	8.5	12.2	10	20.9	13.1	16.0
200	10	18.2	14.0	16.1	10	21.6	18.7	20.1	10	29.2	20.8	23.0

In comparing our results with previous work, the only other computational experience we found was that of Ingargiola and Korsh [15] in which they solve a set of ten problems consisting of 15 objects and 6 knapsacks (occupancy ratio = 0.2). Their average computation time for this set of problems was 7 seconds on an IBM 370-155. We were unable to generate random problems having the characteristics defined by Ingargiola and Korsh, due to our restriction that all item weights be less than or equal to the largest box capacity. When we suspended this restriction, our Lagrangean relaxation procedure required an average of 0.2 seconds per problem.

BIBLIOGRAPHY

[1] Ahrens, J.H., and G. Finke, "Merging and Sorting Applied to the 0-1 Knapsack Problem," Operations Research 23, 1099-1109 (1975).

[2] Balas, E., "Discrete Programming by the Filter Method," Operations Research 19, 915-957 (1967).

[3] Dantzig, G.B., "Discrete-Variable Extremum Problems," Operations Research 5, 266-277 (1957).

[4] Eilon, S., and N. Christofides, "The Loading Problem," Management Science 17, 259-268 (1971).

[5] Fisk, John, "Multiple Knapsack Algorithms," unpublished Doctoral Dissertation, Kent State University, Kent, Ohio, (1974).

[6] Geoffrion, A., "An Improved Implicit Enumeration Approach for Integer Programming," Operations Research 17, 437-454 (1969).

[7] Geoffrion, A., "Lagrangean Relaxation for Integer Programming," Mathematical Programming Study 2, 82-114 (1974).

[8] Geoffrion, A., and R. Marsten, "Integer Programming Algorithms: A Framework and State of the Art Survey," Management Science 18, 465-491 (1972).

[9] Gilmore, P.L., and R. Gomory, "A Linear Programming Approach to the Cutting Stock Problem, Part II," Operations Research 11, 863-888 (1963).

[10] Glover, F., "Surrogate Constraints," Operations Research 16, 741-749 (1968).

[11] Glover, F., "Surrogate Constraint Duality in Mathematical Programming," Operations Research 23, 434-451 (1975).

[12] Greenberg, H., "The Generalized Penalty Function Surrogate Model," Operations Research 21, 162-178 (1973).

[13] Greenberg, H., and W. Pierskalla, "Surrogate Mathematical Programs," Operations Research 18, 924-939 (1970).

[14] Hung, M.S., and J.R. Brown, "An Algorithm for A Class of Loading Problems," Naval Research Logistics Quarterly 25, 289-297 (1979).

[15] Ingargiola, G., and J. Korsh, "An Algorithm for the Solution of 0-1 Loading Problems," Operations Research 23, 1110-1119 (1975).

[16] Pierce, J., and J.S. Lasky, "Improved Combinatorial Programming Algorithms for a Class of 0-1 Integer Programming Problems," Management Science 19, 528-543 (1973).

[17] Ross, G.T., and R. Soland, "A Branch and Bound Algorithm for the Generalized Assignment Problem," Mathematical Programming 8, 91-103 (1975).

[18] Salkin, H.M., and C.A. DeKluyver, "The Knapsack Problem: A Survey," Naval Research Logistics Quarterly 22, 127-144 (1975).

INFORMATION FOR CONTRIBUTORS

The NAVAL RESEARCH LOGISTICS QUARTERLY is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Manuscripts and other items for publication should be sent to The Managing Editor, NAVAL RESEARCH LOGISTICS QUARTERLY, Office of Naval Research, Arlington, Va. 22217. Each manuscript which is considered to be suitable material for the QUARTERLY is sent to one or more referees.

Manuscripts submitted for publication should be typewritten, double-spaced, and the author should retain a copy. Refereeing may be expedited if an extra copy of the manuscript is submitted with the original.

A short abstract (not over 400 words) should accompany each manuscript. This will appear at the head of the published paper in the QUARTERLY.

There is no authorization for compensation to authors for papers which have been accepted for publication. Authors will receive 250 reprints of their published papers.

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics, for possible publication in the NEWS AND MEMORANDA or NOTES sections of the QUARTERLY.

NAVAL RESEARCH
LOGISTICS
QUARTERLY

SEPTEMBER 1978
VOL. 25, NO. 3
NAVSO P-1278

Partial CONTENTS:

OFFICE OF NAVAL RESEARCH
Arlington, Va. 22217